Comparative Molecular Genetics of Humans and Chimpanzees

by

CHERYL ANNE WISE

A thesis submitted for the degree of
Doctor of Philosophy of The Australian National University

John Curtin School of Medical Research
Australian National University
Canberra, Australia

May, 1998

STATEMENT

This thesis describes the results of research undertaken in the Human Genetics Group of the John Curtin School of Medical Research, Australian National University, Canberra. The research was undertaken between January 1994 and December 1997 while I was the recipient of an Australian Postgraduate Research Award.

Except where otherwise stated in the text or acknowledgments, the work presented in this thesis is entirely my own accomplished under the supervision of Dr Simon Easteal.

Cheryl A. Wise

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr Simon Easteal, for his invaluable support, advice and guidance during the course of this research. His faith in my ability kept me motivated when the going got rough.

I have received a great deal of support of various kinds from many people. I have enjoyed scientific discussions with, or the advice of: Drs Alan Thorne, Xiaojiang Gao, Gareth Chelvanayagam, Lars Jermiin, David Betty, Ingrid Jakobsen and Colin Groves.

I would like to acknowledge the more technical assistance and advice of Ms Michaela Sraml and Mrs Lynn Croft, who contributed to the *ND2* sequencing and the tandem repeat loci characterisation. I also wish to thank the Biomolecular Resource Facility at JCSMR for providing a valuable, friendly service.

I am grateful to Dr D. C. Rubinsztein, Department of Medical Genetics, Addenbrooke's Hospital, Cambridge, U.K. for providing tandem repeat loci data, and to Drs D. C. Gajdusek and C. J. Gibbs Jr. for supplying the chimpanzee blood samples.

A big thankyou must go to all members, past and present, of the Human and Molecular Genetics Groups who have provided much assistance and friendship over the past four years. In particular, thanks to Sarah Weenink and Genevieve Herbert for always being there. Thanks also to Peta Joyce and my friends down at The Station for keeping me sane.

Finally, special thanks must go to my family who have emotionally and financially supported me over the years. Without their help I would not be where I am today.

ABSTRACT

For the past decade, mitochondrial DNA (mtDNA) has dominated studies of human genetic variation and evolution. Patterns of human mtDNA variation suggest that all modern humans derive from a common ancestral population present in Africa some 100,000 to 200,000 years ago. This explanation is consistent with some paleontological interpretations, but not with others that imply evolutionary continuity among Old World populations during the past million years. It assumes selective neutrality of the mitochondrial genome and that the patterns of variation reflect the species' history of population size and structure. An alternative explanation is that variation in the human mitochondrial genome is the result of natural selection.

The presence of natural selection provides an interesting alternative hypothesis to be tested. One method is to compare levels of mitochondrial and nuclear genome diversity between closely related species. Limited restriction mapping and sequencing studies have shown that humans have lower levels of mitochondrial genome diversity than chimpanzees. In contrast, humans appear to have higher levels of nuclear genome diversity based on electrophoretically detectable protein polymorphims. Here the discrepancy is investigated further by examining mitochondrial control region sequences and levels of heterozygosity at short tandem repeat (STR) loci, particularly for chimpanzees. Comparison of these data with equivalent published data confirms that humans have a substantially lower ratio of mitochondrial to nuclear genome diversity than chimpanzees and other catarrhines. This implies something unusual in the dynamics of human evolution.

An additional method is to compare patterns of genetic variation with neutral evolutionary models. In this study, nucleotide sequences were determined for the mitochondrial NADH dehydrogenase subunit 2 (*ND2*) gene and the cytochrome b (*cyt b*) gene in humans and chimpanzees. Statistical tests employing various measures of genetic variation were used to investigate neutral predictions. The combined results are consistent with recent directional selection or background selection against slightly deleterious mutations in the human mitochondrial genome.

These results suggest that evolutionary inferences based on mitochondrial variation patterns should be carefully reconsidered. The high degree of genetic similarity between mitochondrial genomes in modern humans may not reflect a recent common exodus from Africa, but the action of natural selection.

TABLE OF CONTENTS

ABBREVIATIONS

ATP	adenosine triphosphate
bp	base pair
DHPLC	denaturing high-performance liquid chromatography
dNTP	deoxynucleotide triphosphate
DNA	deoxyribonucleic acid
ddH_2O	double distilled water
EDTA	ethylene diamine tetra-acetic acid
FET	Fisher's exact test
kb	kilobase
MHC	major histocompatibility complex
Ma	million years
mt	mitochondrial
OD	optical density
OXPHOS	oxidative phosphorylation
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
rpm	revolutions per minute
RNA	ribonucleic acid
STR	short tandem repeat
SSCP	single-strand conformation polymorphism
SDS	sodium dodecyl sulfate
TEMED	tetramethylethylene diamine
ka	thousand years
Tris	trishydroxymethylaminomethane
UV	ultraviolet
VNTR	variable number tandem repeat
V	volt
w/v	weight per volume
YAP	Y *Alu* polymorphism

PUBLICATIONS

Wise, C. A., M. Sraml, D. C. Rubinsztein and S. Easteal. 1997. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. *Mol. Biol. Evol.* **14:** 707-716.

Wise, C. A., M. Sraml and S. Easteal. 1998. Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. *Genetics* **148:** 409-421.

CONFERENCE PROCEEDINGS

Wise, C. A., M. Sraml and S. Easteal. 1994. Out of Anywhere?: Evolution of human mitochondrial DNA does not reflect the evolutionary history of humans; evidence from a comparative analysis of chimpanzees. Genetics Society of Australia 41st Annual Meeting, University of New England, Armidale, Australia.

Wise, C. A., M. Sraml and S. Easteal. 1994. The origin of modern humans: a comparative molecular genetic analysis of humans and chimpanzees. Robertson Symposium, DNA Profiling and Gene Populations, Australian National University, Canberra, Australia.

Wise, C. A., M. Sraml and S. Easteal. 1995. Recent coalescence of human mitochondrial genomes results from natural selection and not the movement of people out of Africa. Genetics Society of Australia 42nd Annual Meeting, Australian National University, Canberra, Australia.

Wise, C. A., M. Sraml and S. Easteal. 1995. Natural selection and the recent coalescence of human mitochondrial genomes. The third international meeting of the Society for Molecular Biology and Evolution, Hayama, Japan.

Wise, C. A., M. Sraml and S. Easteal. 1996. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 2 gene in humans and chimpanzees. 4th International DNA Fingerprinting Conference, Melbourne, Australia.

1.1 THE CONFLICT OVER MODERN HUMAN ORIGINS

The earliest known hominid, *Ardipithecus ramidus*, lived in Africa around 4.4 million years (Ma) ago, but it is uncertain whether it was bipedal or in the direct line of descent to modern humans (White *et al.* 1994, 1995; WoldeGabriel *et al.* 1994). The recently described *Australopithecus anamensis*, dated at 3.9 to 4.2 Ma, was bipedal and has been placed in the line of descent to modern humans (Leakey *et al.* 1995). The first hominid species to leave Africa was *Homo erectus* (now called "early *Homo sapiens*" by many). New fossils from Georgia (Gabunia and Vekua 1995), and older fossils from Java which have been newly dated (Swisher *et al.* 1994), put the time of *H. erectus*' emergence out of Africa between 1.6 and 1.8 Ma ago. This first emergence of hominids out of Africa in the form of so-called *H. erectus* is universally accepted by paleoanthropologists and not under current debate. When and where the transition from archaic humans to modern *H. sapiens* occurred, however, remains controversial.

There are two main hypotheses put forward to explain the origin of modern humans (Figure 1.1). The polycentric theory of human evolution (Weidenreich 1940, 1946) is the intellectual precursor of the multiregional model (*e.g.*, Wolpoff *et al.* 1984; Wolpoff 1989; Frayer *et al.* 1993; Wolpoff and Caspari 1997). This model proposes that throughout the past two million years humans have been a single widespread polytypic species, with multiple, gradually evolving interconnected populations that retained some regional differences (Wolpoff and Caspari 1997). Weidenreich's polycentric theory has often been misrepresented as a candelabra (parallel evolution) by omitting the interconnections (*e.g.*, Howells 1942, 1959, 1993; Coon 1962; Lewin 1993). The alternative (replacement) model of human evolution proposes a recent single origin for modern humans, with worldwide replacement of archaic populations (*e.g.*, Protsch 1975; Howells 1976; Stringer and Andrews 1988; Stringer 1990).

Studies of genetic variation are potentially able to address the questions of when and where modern humans arose. For example, molecular data from mitochondrial DNA (mtDNA) (Cann *et al.* 1987*a*; Vigilant *et al.* 1991) has suggested that modern humans originated in Africa approximately 100 to 200 thousand years (ka) ago (see Section 1.3).

For some researchers (Ruvolo *et al.* 1993; Stoneking 1993), "when" is more important than "where" for deciding among the theories, since an African origin is compatible with all of them.

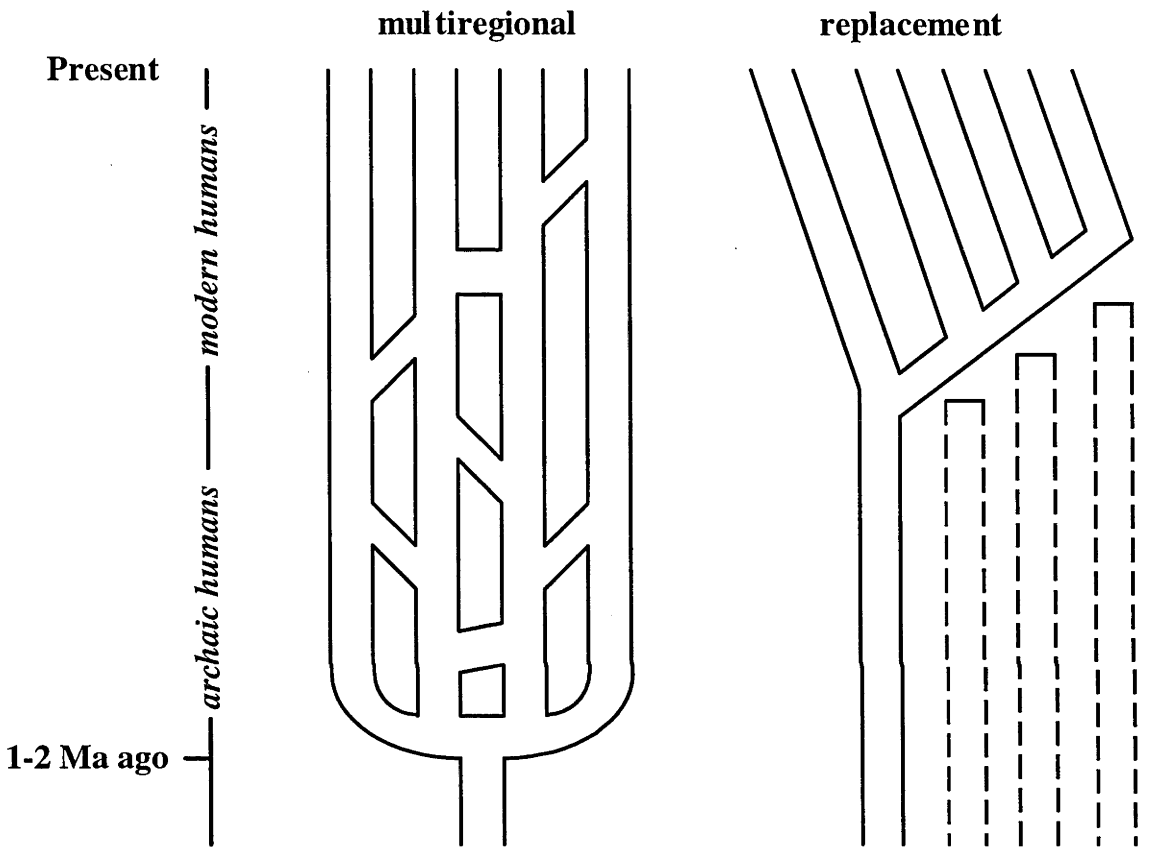


FIGURE 1.1 Two models for the origin of modern humans, modified from Takahata (1993*a*). The multiregional model assumes that human populations were strongly interconnected throughout the Pleistocene. By contrast, the replacement model assumes the complete replacement of archaic humans by modern *H. sapiens*. An African origin at 1–2 Ma versus an African origin at 100–200 ka is the major difference between the multiregional and replacement models.

1.2 PROPERTIES OF MITOCHONDRIAL DNA

Mitochondria are semiautonomous and have the capacity to replicate, divide and fuse independent of the somatic nuclear division cycle (Ankel-Simons and Cummins 1996). In eukaryotic organisms, mitochondria play a principle role in energy production through oxidative phosphorylation (OXPHOS). Most of the proteins found within the mitochondrion are encoded by nuclear DNA and are imported into the organelle from the cytosol after synthesis on cytosolic ribosomes. A small proportion of mitochondrial proteins are, however, encoded by mitochondrial DNA (mtDNA) and synthesised by the separate mitochondrial translation system (Anderson *et al.* 1981).

The human mitochondrial genome is a closed circular, double-stranded molecule consisting of 16,569 base pairs (bp), with the two strands containing either a disproportionate number of guanine (heavy or H-strand) or cytosine (light or L-strand) nucleotide bases (Anderson *et al.* 1981). It encodes 13 proteins involved in OXPHOS, 22 transfer RNAs (tRNA) and two ribosomal RNAs (rRNA) (Figure 1.2). The proteins include two ATP synthetase subunits, three cytochrome oxidase subunits, cytochrome *b*, and seven NADH dehydrogenase subunits (Anderson *et al.* 1981; Attardi *et al.* 1986). Mammalian mtDNA is characterised by having few noncoding bases (Attardi 1985). The control region which has no coding function is the exception with a length of approximately 1,100 bp (Aquadro and Greenberg 1983). It contains the origin of transcription for both strands (Cantatore and Attardi 1980) and the origin of replication for the H-strand (Anderson *et al.* 1981). This region is the most polymorphic region of the genome (Cann and Wilson 1983), with most of the variation distributed not at random but rather concentrated in two hypervariable segments: *segment I* (positions 16024–16401 in the numbering system of Anderson *et al.* 1981) and *segment II* (positions 29–408) (Vigilant *et al.* 1989).

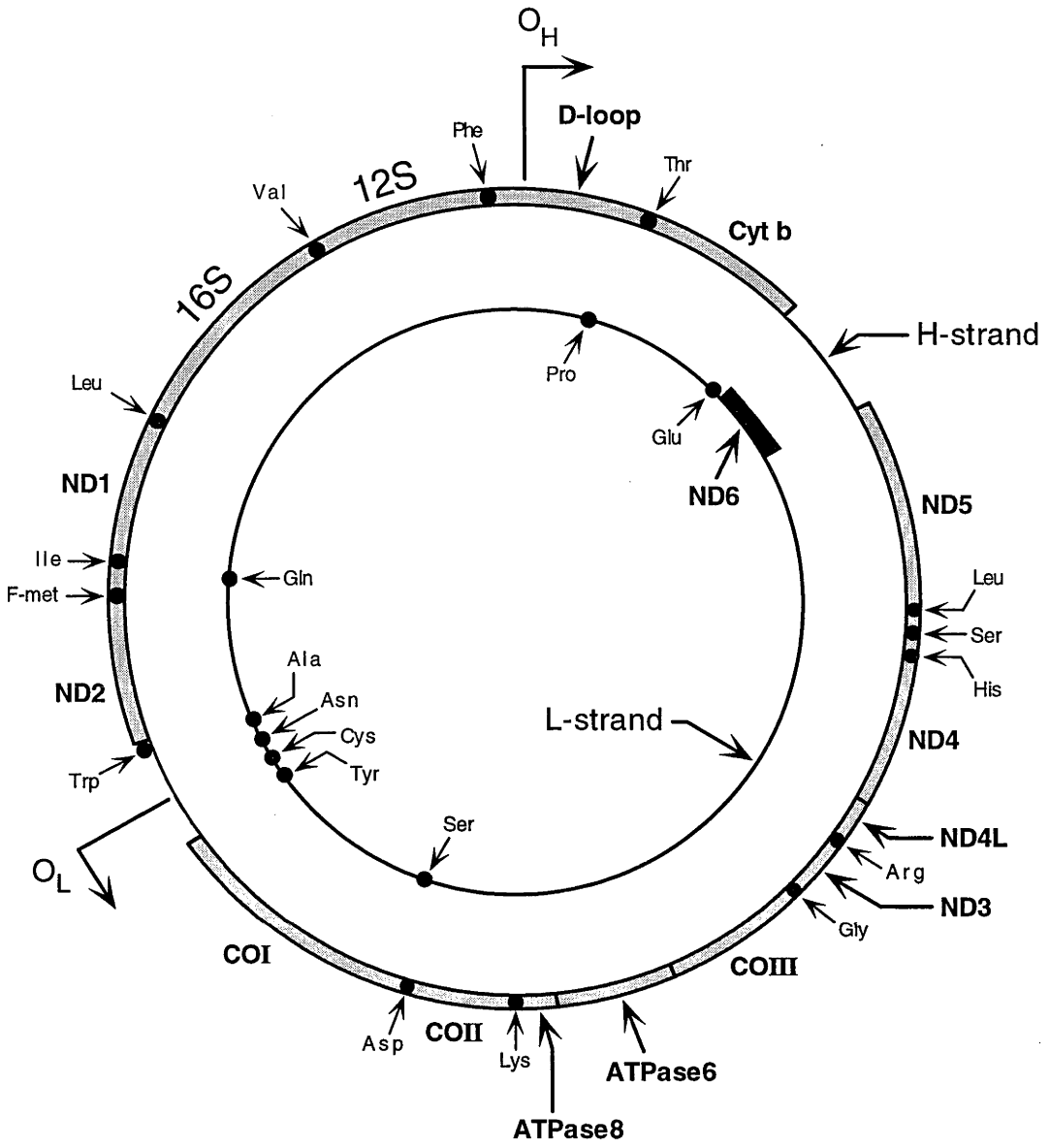


FIGURE 1.2 Genetic organisation of the human mitochondrial genome. The outer circle indicates those genes transcribed from the H (heavy)-strand and the inner circle indicates those genes transcribed from the L (light)-strand. The complete genome is 16,569 base pairs in length and contains two ribosomal RNA genes (12S and 16S), 22 transfer RNA genes (designated by the three letter amino acid code) and 13 protein-coding genes. ATPase 6 and 8 are subunits of the mitochondrial ATPase complex. COI, COII and COIII are cytochrome c oxidase subunits. Cyt b stands for cytochrome b. ND1 to ND6 are subunits of the NADH dehydrogenase complex. O_H and O_L are the origins of replication of the H-strand and L-strand respectively; the directions of replication are shown by the arrows. The major noncoding region between the Phe- and Pro-tRNA genes is called the D-loop or control region.

Mitochondrial DNA possesses a number of advantages over nuclear DNA which make it particularly applicable in the study of such fields as molecular evolution, population genetics, palaeontology and forensic science. These include predominantly maternal inheritance, no recombination, homoplasmy, multiple copy number and high mutation rate (Wilson *et al.* 1985; Avise 1986).

1.2.1 Maternal Inheritance and Homoplasmy

Maternal inheritance of mtDNA has been observed in a number of vertebrates including *Xenopus* (Dawid 1972), horse-donkey hybrids (Hutchison *et al.* 1974), birds (Watanabe *et al.* 1985), mice (Gyllensten *et al.* 1985) and humans (Giles *et al.* 1980). The simplest explanation for so-called maternal inheritance is that the paternal contribution is diluted to almost undetectable levels (Ankel-Simons and Cummins 1996). The typical mammalian sperm midpiece contains approximately 50 to 75 mtDNA molecules. In contrast, the mammalian oocyte contains around 10^5 to 10^8 copies of mtDNA. Thus, the oocyte's mtDNA copy number exceeds that of the sperm by a factor of at least 10^3 (Ankel-Simons and Cummins 1996). Heteroplasmic cells are rarely found (Avise *et al.* 1987; Avise 1991) and heteroplasmy appears to be more prevalent in lower vertebrates and invertebrates than in birds or mammals (Bermingham *et al.* 1986). Since individuals are predominantly haploid and homoplasmic, a straight-forward mode of genetic transmission is exhibited for mtDNA without the segregation and recombination associated with nuclear DNA during sexual reproduction.

1.2.2 High Copy Number

Eukaryotic cells contain between 10^2 to 10^4 mitochondria, with each organelle carrying 4 to 6 mitochondrial genomes (Birky 1983). Thus, any one cell might normally contain thousands of copies of mtDNA, compared with only two copies per cell of nuclear DNA. The high copy number confers the advantage of requiring a sample a fraction of the size required for traditional nuclear DNA analysis. This factor is of particular relevance in conservation and forensic contexts where sample size and quality may be limited (*e.g.*, Pääbo 1989; Stoneking *et al.* 1991; Sullivan *et al.* 1992; Holland *et al.* 1993).

1.2.3 High Mutation Rate

Mitochondrial DNA has been reported as having a mutation rate five to ten times faster than nuclear DNA (Brown *et al.* 1979), based on restriction endonuclease and thermo-stability analyses between closely related primate species. Single-copy nuclear DNA showed low levels of divergence while mtDNA showed extensive divergence, implying a more rapid rate of evolution in the mtDNA.

Increased mutation rates for the mtDNA molecule are thought to be the result of general factors such as error prone replication, inefficient repair, rapid DNA turnover and/or high oxidant levels (*e.g.*, Brown *et al.* 1979, 1982; Brown 1983). Mitochondrial DNA is replicated by γ-polymerase, an enzyme that has a higher error rate than the major nuclear DNA replication enzyme, α-polymerase (Kunkel and Loeb 1981). Furthermore, attempts to measure nucleotide excision repair in mitochondria failed to detect activity (Clayton *et al.* 1974), prompting speculation that mammalian mitochondria lack DNA repair (Clayton 1982; Miquel 1992). However, recent studies suggest that while mammalian mitochondria lack nucleotide excision repair activity, they may possess other DNA repair pathways (*e.g.*, Thyagarajan *et al.* 1996). Mitochondrial DNA also has a short turnover time with many more rounds of replication than nuclear DNA, whose generation time is identical to the cell type in which it occurs. Consequently, mtDNA may have an increased number of mutations produced per cell generation as a result of replication errors (Brown 1983). Mitochondrial DNA is also particularly susceptible to mutation due to its close proximity to free radicals generated by OXPHOS and its lack of protective histones (Shoffner and Wallace 1992).

An additional factor that could be responsible for the high rate of mtDNA evolution is relaxed translational constraints (Brown *et al.* 1982; Cann *et al.* 1984; Wilson *et al.* 1985). The translation apparatus of a small genome, such as mtDNA, which encodes only 13 polypeptides, would be expected to function under more relaxed constraints than would a system that translates thousands of different mRNAs (Cann *et al.* 1984). Consistent with this view is the observation that the tRNA and rRNA genes of mtDNA evolve 100 times faster than their nuclear counterparts (Brown *et al.* 1982; Cann *et al.* 1984).

The divergence rate of mitochondrial genomes has been estimated as 2–4% per Ma, that is, 2 to 4 substitutions per hundred base pairs per million years (Brown *et al.* 1979; Wilson *et al.* 1985; Stoneking *et al.* 1986). These estimates are derived from interspecies comparisons of diverse vertebrate species, together with intraspecies comparisons in humans. The estimated rate for the noncoding control region is approximately ten times faster than the genome as a whole (Aquadro and Greenberg 1983; Vigilant *et al.* 1989, 1991; Horai *et al.* 1995). Such a high rate of evolution makes mtDNA highly attractive for the study of closely related species and populations within species. The maternal, haploid inheritance of mtDNA means that, with no genetic recombination, the only source of new variation is mutation. Therefore, the number of mutations separating two mtDNA types reflects how closely related they are—the larger the number of mutations, the more distantly related the mtDNA types. Phylogenetic trees can therefore be readily constructed and interpreted as reflecting the maternal genealogical history of a population or species. Time scales can be affixed to these trees by determining both the amount of sequence divergence that has accumulated since any two mtDNA types last shared a common ancestor, and the rate of mtDNA evolution (Stoneking 1993).

It is important to note, however, that there are two different ways of expressing the amount of evolutionary change. As Figure 1.3 illustrates, one can measure either the number of mutations that have occurred from a mtDNA type back to an ancestor or the number of mutations that have occurred between two mtDNA types. The first measure is the amount of *evolution* or change along one lineage, the second measure is the amount of *divergence* or change between two lineages. Translated into time, the rate of divergence is expected to be twice the rate of evolution because it incorporates changes in two lineages (Stoneking 1993).

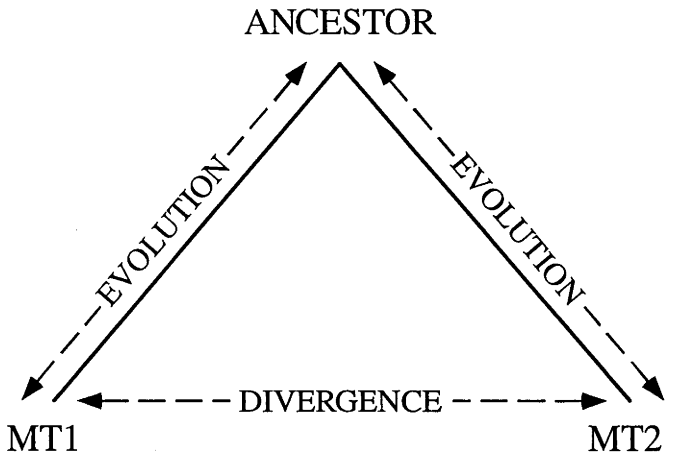


FIGURE 1.3 An illustration of the principle that the rate of sequence divergence is twice the rate of sequence evolution, taken from Stoneking (1993). For two contemporary mtDNA types (MT1 and MT2), the amount of evolution is the number of sequence changes from the ancestor to either mtDNA type, whereas the amount of divergence is the number of sequence changes between the two mtDNA types. Hence, the amount (or rate) of divergence is twice the amount (or rate) of evolution.

1.3 USING MITOCHONDRIAL DNA TO STUDY MODERN HUMAN ORIGINS

Two major means of studying human mtDNA variation include restriction mapping and nucleotide sequencing.

1.3.1 Restriction Mapping

Early human mitochondrial-based evolutionary studies utilised restriction endonuclease analyses. Restriction enzymes recognise either four (Brown 1980) or six-base (Johnson *et al.* 1983) sequences (usually an inverted repeat) and cleave the DNA in or adjacent to the recognition sequence. The presence or absence of restriction sites produces various length variants which is a simple way to detect polymorphism either within or between populations or species (Wilson *et al.* 1985). Length mutations can also be detected by these methods (Cann and Wilson 1983).

The most notable of the restriction mapping studies was performed by Cann *et al.* (1987*a*), who used 12 restriction enzymes to map 147 individuals from five geographic regions. From these individuals they found 133 distinct mtDNA types which were used to construct a genealogical tree by maximum parsimony and midpoint rooting methods. The resulting tree showed two major clades, one consisting exclusively of Africans and the other containing individuals from all populations, including Africa. Moreover, African populations exhibited more mtDNA sequence divergence than non-African populations. From these observations, Cann *et al.* (1987*a*) concluded that the common ancestor of all modern human mtDNA types probably lived in Africa. Given a sequence divergence of 0.57%, and assuming the rate of mtDNA sequence divergence to be 2–4% per Ma (Stoneking *et al.* 1986), it was calculated that the common mtDNA ancestor existed 140,000 to 290,000 years ago. Another global analysis of mtDNA restriction-site variation revealed that African mtDNAs are approximately 101,000 to 133,000 years old and that African mtDNA diversity is the greatest of all populations (Chen *et al.* 1995). Furthermore, non-African haplotypes appear to be derived from a subset of African mtDNAs. Other restriction mapping studies have also found that, on average, African populations have the greatest mtDNA diversity, followed by Asian and Caucasian populations (*e.g.*, Johnson *et al.* 1983; Excoffier and Langaney 1989; Excoffier 1990; Merriwether *et al.* 1991). These observations have been interpreted as support for the replacement theory, since multiregional evolution predicts that molecular divergence times should be upwards of one million years old.

Criticisms of the Cann *et al.* (1987*a*) study centred on the midpoint rooting method used to locate the mtDNA ancestor on the tree (Darlu and Tassy 1987). This is an indirect method that assumes a constant rate of evolution. In general, it is preferable to use a direct method, such as outgroup rooting, by including a closely related non-human mtDNA type in the analysis. Other criticisms include the use of a small sample of African Americans to represent native African mtDNAs (Darlu and Tassy 1987), the mathematical theory used for estimating the age of the mtDNA ancestor (Krüger and Vogel 1989), and the methods used for calibrating the rate of human mtDNA evolution (Saitou and Omoto 1987). The latter argue that the rate is 0.7% per Ma, or about half that of previous estimates (2–4%) in primates (Brown *et al.* 1979). However, as pointed out by

Cann *et al.* (1987*b*), the 2–4% rate refers to the comparison of two lineages, whilst the 0.7% rate refers to the amount of evolution along one lineage (refer Figure 1.3). The strongest criticism has, however, resulted from the re-analysis by Maddison (1991). He found that maximum parsimony trees requiring fewer mutations could be constructed from the data, and that at least some of these trees did not have an African origin.

1.3.2 Mitochondrial DNA Sequence Data

Sequence analysis has the potential to detect all mtDNA mutations and thus provides the maximum degree of resolution. It enables the comparison of specific regions of DNA between closely related individuals or populations to obtain divergence estimates (Wilson *et al.* 1985). With the advent of the polymerase chain reaction (PCR), restriction mapping and sequencing via conventional cloning have been replaced by enzymatic amplification and direct sequencing (*e.g.*, Wrischnik *et al.* 1987; Gyllensten and Erlich 1988; Gyllensten 1989).

1.3.2.1 Hypervariable control region sequences

Sequencing studies of the mitochondrial control region have revealed several important features of mtDNA variation, most notably a strong bias favouring transitional over transversional nucleotide changes and an abundance of insertions and deletions (*e.g.*, Aquadro and Greenberg 1983; Greenberg *et al.* 1983; Vigilant *et al.* 1989; Horai and Hayasaka 1990; Kocher and Wilson 1991; Horai *et al.* 1993; Mountain *et al.* 1995; Watson *et al.* 1996). These findings demonstrate the importance of resolution at the nucleotide level, as conferred by direct sequencing.

Vigilant *et al.* (1989, 1991) sought to address the criticisms of the original Cann study by sequencing the two hypervariable segments of the mitochondrial control region which evolve several times faster than other regions of the genome (Aquadro and Greenberg 1983). Sequence analysis of this region therefore offers a high degree of resolution that discriminates among even very closely related individuals. Additional African groups were surveyed and a common chimpanzee sequence was used to root the resulting parsimony trees. The results seemed to confirm the original Cann *et al.* (1987*a*) study in

indicating both a tree topology supportive of an African origin and a recent divergence time. Furthermore, Vigilant *et al.* (1991) applied two statistical tests to their set of parsimony trees and concluded that there was statistical support for an African origin.

The conclusions of Vigilant *et al.* (1991) have been challenged by subsequent reanalyses that found the original phylogenetic analysis to be inadequate (Hedges *et al.* 1992; Maddison *et al.* 1992; Templeton 1992, 1993). Another problem was that the chimpanzee outgroup is so distantly related to humans that multiple substitutions have affected many positions, making it difficult to determine how chimpanzee mitochondrial sequences relate to human sequences (von Haeseler *et al.* 1995). However, new methods of phylogenetic analysis have continued to support an African origin of human mtDNA (Penny *et al.* 1995; D'Andrade and Morin 1996; Strimmer and von Haeseler 1997), as has the use of a nuclear mtDNA insertion (Zischler *et al.* 1995) or the recently described Neanderthal mtDNA sequence (Krings *et al.* 1997) as an outgroup.

Other sequencing studies of the control region have also found that African populations, on average, exhibit more mtDNA diversity than non-African populations (*e.g.*, Horai and Hayasaka 1990; Horai *et al.* 1993; Graven *et al.* 1995; Jorde *et al.* 1995; Watson *et al.* 1996). This pattern of genetic variation has commonly been interpreted as supporting a model in which an (older) African population gave rise to founder population(s) that carried subsets of genetic diversity to Asia and Europe (Relethford 1995). An important assumption of this model is that the higher diversity in Africa reflects a larger accumulation of mutations, and therefore greater time depth. However, as pointed out by several authors, divergence time is not the only factor than can affect levels of variation (*e.g.*, Templeton 1993; Relethford and Harpending 1994, 1995; Relethford 1995, 1997; Rogers and Jorde 1995; Harpending *et al.* 1996). Other influencing factors include mutation rates, effective population size, population subdivision, migration patterns and selection.

Irrespective of geographical origin, the mtDNA data can be used to estimate the age of the most recent common ancestor by reference to an assumed molecular clock. Phylogenetic analysis is the most commonly used approach for estimating the rate of human mtDNA evolution. The mtDNA nucleotide sequences of a sample of individuals are used to construct a phylogenetic tree. The number of mutations that have occurred along each branch of the tree is then estimated. Finally, some independent assumption about the age of the tree (*e.g.*, the time of the human-chimpanzee divergence) is applied to derive a rate of divergence. Using this approach, Vigilant *et al.* (1991) estimated a control region substitution rate, with an attempt to correct for multiple substitutions at the same site. Assuming a human-chimpanzee divergence time of 4–6 Ma, the age of the human mtDNA ancestor was placed between 166,000 and 249,000 years ago. The range reflects only the uncertainty in the calibration rate and does not include the standard error. A variety of methods that accommodate various rates and classes of substitution have been developed for estimating the age of the mtDNA ancestor and the associated standard errors (*e.g.*, Hasegawa and Horai 1991; Nei 1992; Pesole *et al.* 1992; Stoneking *et al.* 1992; Hasegawa *et al.* 1993; Tamura and Nei 1993; Wills 1995). The results of applying these methods to various human mtDNA datasets are summarised in Table 1.1 in the form of approximate 95% confidence intervals for the age of the human mtDNA ancestor. Despite the diverse methods and datasets used, almost all of these confidence intervals place the upper limit for the age of the most recent common ancestor at about 500,000 years ago. However, it is important to note that these estimates may not be directly comparable as they have been calculated using different correction methods and different transition:transversion ratios.

TABLE 1.1 Estimates of the age of the human mtDNA ancestor and associated 95% confidence intervals.

Study	Ancestor age	95% C.I.
Hasegawa and Horai (1991)	280,000	180,000–380,000
Nei (1992)	207,000	110,000–504,000
Pesole *et al.* (1992)	400,000	200,000–600,000
Stoneking *et al.* (1992)[a]	133,000	63,000–356,000
Stoneking *et al.* (1992)[b]	137,000	63,000–416,000
Hasegawa *et al.* (1993)	211,000	0–433,000
Tamura and Nei (1993)	160,000	80,000–480,000
Wills (1995)	436,000	336,000–481,000

NOTE.—All ages and confidence intervals are based on a human-chimpanzee mtDNA divergence date of 4 to 6 Ma ago, with the exception of Stoneking *et al.* (1992). This study used mtDNA divergence specific to Papua New Guinea (PNG) and an initial colonisation date for New Guinea of 60,000 years ago to calibrate the rate of human mtDNA evolution.

[a] Based on the average distance to the root of each PNG group.

[b] Based on the corrected divergence between PNG groups.

The degree of correction for the control region data is large, and therefore worrisome in terms of the accuracy of inferred molecular divergence dates. The observed amount of sequence divergence between human and chimpanzee mtDNA must be corrected for multiple substitutions at the same site. The dynamics of mtDNA sequence evolution (rapid rate of change, elevated frequency of transition-type mutations, and existence of mutational "hotspots") makes this correction particularly troublesome. The difficulty is confounded by the relatively long period of time separating humans and chimpanzees (Stoneking 1993). Wills (1995) attempts to avoid these problems by estimating the age of the mtDNA ancestor from much lower frequency transversion mutations. Assuming a linear accumulation of transversions, a chimpanzee-human split ranging between 4.0–7.4 Ma ago gives a coalescence age of 336,000 to 889,000 years ago.

A more recent problem comes from intergenerational calculations of control region mutation rates which are much higher than estimates derived from phylogenetic studies (Howell *et al.* 1996; Parsons *et al.* 1997; Parsons and Holland 1998). Using the intergenerational mutation rate to calibrate the mtDNA molecular clock would result in an age of only 6,500 years ago. However, it is unlikely that the current distribution of human populations and genes has arisen in the past few hundred thousand years (Loewe and Scherer 1997). Thus, the study of more slowly evolving protein coding mitochondrial genes may prove more useful. Less correction is needed as fewer differences are expected among sequences. Therefore, estimated divergence times for protein coding mitochondrial genes are potentially more accurate than those provided by the noncoding control region (Ruvolo 1996).

1.3.2.2 Mitochondrial protein coding sequences

The hypervariable segments of the mitochondrial control region evolve approximately ten times faster than other regions of the genome (Aquadro and Greenberg 1983). Furthermore, control region sequences clearly display heterogeneity among sites (Kocher and Wilson 1991). The implication is that multiple substitutions at the same site accumulate rapidly in this region. The slower rate of evolution for the protein coding genes means that fewer differences are observed between humans and chimpanzees, compared with the control region. Thus, these regions are less likely to become saturated with multiple substitutions and an appropriate correction can be made.

Kocher and Wilson (1991) sequenced an 896 bp segment which includes parts of the genes for NADH dehydrogenase subunits 4 and 5 (*ND4-5*). The rate of heterogeneity among sites of third codon positions seems to be less serious than that of noncoding regions. Assuming a human-chimpanzee divergence of 5 Ma, they estimated that the human mtDNA ancestor lived approximately 170,000 years ago, but did not provide a standard error for this estimate. Ruvolo *et al.* (1993) combined data from the slowly evolving mitochondrial cytochrome oxidase subunit II (*COII*) gene with that from the *ND4-5* region. Assuming the latest possible human-chimpanzee divergence date to be 6 Ma ago, the combined *COII* and *ND4-5* data place the human mtDNA ancestor at 298,000 years, with a 95% confidence interval of 129,000 to 536,000 years. Thus, the

protein coding mitochondrial data also appear to support a recent origin for modern humans.

1.3.2.3 Complete mitochondrial DNA sequences

Recently, the complete mtDNA sequences of three humans (African, European and Japanese), three African apes (common and pygmy chimpanzees, and gorilla), and one orangutan were reported in an attempt to more accurately estimate the substitution rates and divergence times of hominoid mtDNAs (Horai *et al.* 1995). From these data a human-chimpanzee divergence time of 4.9 Ma was obtained, and the age of the human mtDNA ancestor was estimated as 143,000 years.

The approach of using entire mtDNA sequences to address modern human origins is powerful as it maximises the number of slowly evolving mtDNA sites to be studied. Since these require little data correction, dates derived from them are potentially more accurate. Sequencing complete mitochondrial genomes reveals the maximum possible number of differences among individuals, and should be pursued not only for what it can tell us about modern human origins, but because comparisons of closely related human DNA sequences can help us elucidate molecular evolutionary processes (Ruvolo 1996).

1.3.3 Mismatch Distributions

For several years now the genetic literature on modern human origins has relied heavily on phylogenetic trees inferred from mtDNA data. An alternative method of analysing mtDNA sequences has been developed by Rogers and Harpending (1992) based on "mismatch distributions" (or the distribution of pairwise differences). Mitochondrial mismatch distributions are presented as histograms showing the number of nucleotide (or restriction site) differences between all pairs of individuals. Instead of estimating the time of the most recent common ancestor, they use the data to obtain information about population structure and history (Rogers and Harpending 1992; Harpending *et al.* 1993; Sherry *et al.* 1994; Rogers 1995; Rogers and Jorde 1995). Human mismatch distributions characteristically take the form of "waves", with a peak in the wave near the average number of pairwise differences in each population (*e.g.*, Di Rienzo and Wilson 1991;

Horai *et al.* 1993; Mountain *et al.* 1995; Watson *et al.* 1996). Rogers and Harpending (1992) argue that this pattern is the signature of an ancient population expansion (or bottleneck). The shape and position of the wave provide information about the timing and the magnitude of the expansion or bottleneck (Harpending *et al.* 1993; Rogers 1995). Application of this method to various mtDNA datasets supports a "cultural" or Pleistocene explosion model (*e.g.*, see Noble and Davidson 1991) for a very recent major expansion of the human population, perhaps within the past 40,000 years (Harpending *et al.* 1993; Sherry *et al.* 1994; Rogers 1995).

It is important to note that natural selection is also expected to result in unimodal (wave shaped) mismatch distributions (Slatkin and Hudson 1991; Rogers and Jorde 1995). Thus, human mitochondrial mismatch distributions may reflect an expansion of mtDNA types due to selection, rather than an actual population expansion.

1.3.4 Selection Hypothesis

Estimates of mtDNA diversity in human populations are remarkably low compared to other species (Ferris *et al.* 1981; Wilson *et al.* 1985; Ruvolo *et al.* 1994; Rogers and Jorde 1995; Wise *et al.* 1997). At equilibrium, the diversity of selectively neutral genes is directly proportional to effective population size, N_e (Kimura 1983). A recent population bottleneck in the human lineage has previously been suggested as the cause of low mtDNA diversity in humans compared to other hominoids, whose populations are much smaller (Brown 1980; Wilson *et al.* 1985). Rogers and Jorde (1995) argue that the mismatch waves in the human data reflect changes in population size rather than natural selection. They postulate that the wave in the human data is due to an environmental catastrophe that has also affected other species, such as east African chimpanzees which show a similar wave. Although either wave could have been produced by a favourable mutation sweeping through the population, it is unlikely that this would occur in both species at the same time. Additional analysis of mitochondrial control region sequences from chimpanzees shows multimodal distributions in some chimpanzee populations (see Section 3.4.1; Wise *et al.* 1997), making an explanation in terms of an environmental catastrophe less likely. Furthermore, chimpanzee subspecies are mitochondrially more

diverse than humans, based on sequence data from the noncoding control region (see Section 3.4.1; Morin *et al.* 1994; Wise *et al.* 1997) and from a number of coding genes (see Section 4.4.1; Ruvolo *et al.* 1994; Nachman *et al.* 1996; Wise *et al.* 1998). Limited sequence data from gorillas and orangutans shows that they also exhibit high levels of mtDNA diversity (Ruvolo *et al.* 1994; Garner and Ryder 1996; Xu and Arnason 1996). Thus, whatever caused relatively low mtDNA diversity within humans did not affect other hominoids. Any significant environmental or climatological event affecting human mtDNA should have affected chimpanzees and gorillas, as well as a diverse array of other species. Evidently this is not the case, and it is therefore unlikely that a recent environmental catastrophe produced low genetic diversity within humans. There must be some alternative explanation for the unusual (from a hominoid perspective) mtDNA pattern found within humans.

Another possible explanation is that mtDNA has undergone selection. For example, if a selectively favoured mutation had arisen anywhere in the human mitochondrial genome, then the mtDNA data would reflect the spread and fixation of that advantageous mutation, not human population movements *per se*. Time estimates would refer to the time of this selective sweep of the new mtDNA type through the population. This could lead to a serious misinterpretation of the data, when neutrality is assumed. It is thus important to test the assumption of selective neutrality before interpreting human mtDNA data. This can be done by comparing: (1) Human mitochondrial variation with nuclear variation (see Section 1.5.1). (2) Levels of mitochondrial and nuclear variation between humans and chimpanzees (see Section 1.5.2). (3) Mitochondrial variation patterns with neutral evolutionary models (see Section 1.6).

1.4 NUCLEAR DNA STUDIES

Studies of nuclear DNA have lagged behind mitochondrial DNA for a variety of technical reasons and because processes such as recombination and gene conversion make interpretation of patterns of variation in nuclear DNA more difficult. Nevertheless, it is imperative that we study nuclear DNA. Mitochondrial genomes are only a fraction of the size of nuclear genomes, and the amount of information that they contain is limited.

Furthermore, the pattern of variation at a single genetic locus, such as mtDNA, might, by chance or by selection, differ drastically from other loci and therefore may not accurately reflect historical population processes.

1.4.1 Protein Polymorphisms

Phylogenetic trees constructed from nuclear gene frequency data differ from those used to analyse mtDNA data in that they represent populations, not individual genetic types. A criticism of population trees is that they do not necessarily represent the historical pathways of the populations. They could either represent population divergences or reflect the long-term patterns resulting from gene flow or admixture among populations (Weiss and Maruyama 1976). Nevertheless, population trees are useful representations of gene frequency data as long as one keeps it in mind that they indicate only overall population similarities. Such trees do not indicate the extent to which such similarities are the result of shared evolutionary history rather than migration (Stoneking 1993).

The first studies of genetic variation in humans analysed gene products, such as blood groups, serum proteins, and red cell enzymes (Nei and Roychoudhury 1974*a*, 1982; Cavalli-Sforza *et al.* 1994). Much of the older literature and a thorough analysis of the blood group, tissue antigen, and protein electrophoresis data are summarised in Nei and Roychoudhury (1982). They suggested that Europeans and Asians are genetically closer to each other than to Africans. Nei and Roychoudhury's suggestion was supported by Cavalli-Sforza *et al.* (1988) using 42 polymorphic loci (120 alleles) in 42 worldwide populations. Other studies have also demonstrated that Africans are statistically distinguishable from the other two groups (Nei and Livshits 1989; Nei and Ota 1991). Nei and Roychoudhury (1993) investigated 29 polymorphic loci (121 alleles) in 26 worldwide populations and observed that the first evolutionary splitting of humans occurred between African and non-African populations. This primary split between Africans and all other populations is consistent with either an African origin of modern humans or with much greater gene flow among non-African populations than between African and non-African populations (Stoneking 1993). Hence, the tree topology by itself cannot be used to distinguish between multiregional and single origin (replacement) models.

The time of past population divergences is more critical than geographic origin for distinguishing between the two theories. Both acknowledge that hominids arose in Africa and that the earliest hominid migrations out of Africa probably began about 1–2 Ma ago, involving members of the genus *Homo* (Stringer and Andrews 1988; Wolpoff 1989). Dates derived from protein polymorphism data indicate that Africans diverged from non-Africans about 110,000 to 120,000 years ago (Nei and Roychoudhury 1974*a*, 1982; Nei 1985). Although these divergence times seem to support the recent origin model, they do not take migration into account. This means that these divergence times probably underestimate the actual population divergence times (Weiss and Maruyama 1976; Weiss 1988). A further reason for being cautious about these divergence times is that they depend on uncertain assumptions about rates of amino acid substitution and the proportion of substitutions that are detectable by electrophoresis (Nei 1987).

1.4.2. Restriction Fragment Length Polymorphisms

The study of evolutionary relationships among humans from the major geographical regions using nuclear DNA variation, as assessed by restriction enzymes, was pioneered by Wainscoat *et al.* (1986). These investigators examined five polymorphic restriction sites in a 32 kilobase (kb) region 5′ to the β-globin gene. They analysed 601 chromosomes from eight diverse populations and found that all non-African populations share three common haplotypes whereas African populations have a common haplotype not found in other populations. Genetic distance analysis of haplotype frequencies indicates a major division of human populations into an African and a Eurasian group. Long *et al.* (1990) further analysed the β-globin region in 222 chromosomes from Africa, Eurasia and the Pacific Island. They investigated the phylogenetic relationships among the haplotypes, their presence or absence, and frequencies within populations, and concluded that Africans have more diversity than Europeans or mainland Asians. In a further study of β-globin haplotypes, 852 chromosomes from 12 populations in the Asia-Pacific region were compared with previously published data from other populations, showing some limited support for an African origin of modern humans (Chen *et al.* 1990).

Bowcock *et al.* (1987, 1991*a*) examined 100 restriction fragment length polymorphisms (RFLPs) in five populations and concluded that African populations are the most divergent. Further analysis of the data (Bowcock *et al.* 1991*b*) suggests that Europeans are an admixture of African and Asian elements. Mountain and Cavalli-Sforza (1994) looked at 79 RFLPs in eight human populations and three nonhuman primates (chimpanzees, gorillas and orangutans). Phylogenetic trees based on genetic distances indicated an initial separation between African and non-African populations. Humans share an allele with chimpanzees for 62 polymorphisms. Average frequencies of these ancestral alleles strengthen the conclusion that the earliest major separation of modern humans was between Africans and non-Africans. Jorde *et al.* (1995) examined 30 RFLPs in 243 Africans, Asians and Europeans and also found higher divergence between Africa and the other two populations. However, these studies indicate that Europeans exhibit the highest level of heterozygosity, while Africans have the lowest heterozygosity level. This pattern could reflect the fact that most of these polymorphisms were first detected in European populations (Mountain and Cavalli-Sforza 1994).

Although all of these studies favour a primary split between African and non-African populations they do not provide estimates of the date of separation. Since both the multiregional model and the single origin (replacement) model assume an African origin, the demonstration of an early separation of Africans does not distinguish between them.

1.4.3 Tandem Repeat Loci

Another class of polymorphic loci used to study modern human origins is based on tandemly repeated DNA sequences. Loci with repeated motifs of two to five base pairs are referred to as microsatellites (Litt and Luty 1989) or short tandem repeats (STRs) (Craig *et al.* 1988), while those with longer motifs (*e.g.*, 15 bp) are referred to as minisatellites (Jeffreys *et al.* 1985) or variable number of tandem repeats (VNTRs) (Nakamura *et al.* 1987). Polymorphism is due to variation in the number of tandemly repeated motifs and is considerably greater than the classical serologic and biochemical markers. Such highly polymorphic loci are particularly useful for interpopulation studies, although less useful for comparisons among species (Bowcock *et al.* 1994).

Armour *et al.* (1996) mapped the internal structures of tandem-repetitive alleles from different human populations at the minisatellite locus MS205 (*D16S309*) and found that African populations have the greatest allelic diversity. As noted previously (see Section 1.3.2.1), the simple demonstration of higher diversity in Africa does not necessitate an African origin; for example, greater diversity could result if effective population sizes have historically been higher in Africa. However, the analysis also shows that the lineages present outside of Africa form a subset of the variation existing within Africa. This they claim is most simply consistent with a founder non-African population arising as a subset of a larger African population. However, it is important to remember that this system represents a single locus subject to large stochastic variation.

Edwards *et al.* (1991, 1992) studied genetic differentiation among major racial groups, using five trimeric and tetrameric tandem repeat loci. A comparison of pairwise genetic distances between the three major racial groups reveals a greater divergence for Africans than for the other two groups. Di Rienzo *et al.* (1994) analysed ten dinucleotide repeat loci $(CA)_n$ in Sardinian, Egyptian and sub-Saharan African populations, and came to a conclusion exactly parallel to the above (Africans are more divergent). Bowcock *et al.* (1994) examined 30 microsatellite loci (primarily CA repeats) in 14 human populations and found a significantly higher heterozygosity and number of alleles in Africa compared with other continents. Trees constructed from pairwise genetic distances show that individuals generally cluster according to their geographic origin. This contrasts sharply with trees derived from mtDNA in which individuals of related geographic origin rarely form discrete clusters (Cann *et al.* 1987*a*; Vigilant *et al.* 1991). Deka *et al.* (1995) also examined evolutionary relationships among populations, using microsatellite loci, in terms of genetic distances. They looked at eight dinucleotide repeat loci $(CA)_n$ in eight human populations, and found the highest diversity (high heterozygosity and a large number of alleles) within Africa. Jorde *et al.* (1995) demonstrated greater African heterozygosity in a sample of 30 tetranucleotide repeat loci, but this difference was not statistically significant. In an extension of this study, Jorde *et al.* (1997) analysed an additional 30 microsatellite loci and found that Africans have significantly greater diversity than either Asians or Europeans. A method similar to the mitochondrial mismatch distributions has recently been developed for the analysis of microsatellite data (Shriver

et al. 1997). Applied to three microsatellite datasets these distributions indicate an earlier population expansion and/or a larger effective population size in Africa. This is consistent with an African origin of modern humans, but does not exclude admixture between populations moving out of Africa and local archaic populations (Shriver *et al.* 1997).

A variety of other methods have been used to study microsatellite data (*e.g.*, Bowcock *et al.* 1994; Goldstein *et al.* 1995*a*, 1995*b*; Shriver *et al.* 1995; Slatkin 1995; Zhivotovsky and Feldman 1995; Michalakis and Excoffier 1996), and these incorporate different models of molecular evolutionary change for microsatellite variation. Goldstein *et al.* (1995*a*, 1995*b*) developed an "absolute" dating method based on microsatellite variation. Dating is absolute because the mutation rate for a sample of microsatellite loci has been measured directly in the laboratory from untransformed cells of parents and offspring (Weber and Wong 1993). This method of estimating the mutation rate circumvents the usual reliance on an externally calibrated rate of molecular evolution. Applying this approach to the microsatellite data of Bowcock *et al.* (1994), Goldstein *et al.* (1995*a*) estimated that the deepest split in the human phylogeny occurred between Africans and non-Africans about 156,000 years ago, with a 95% confidence interval of 75,000–287,000 years (where error due to coalescence is included but error associated with mutation rate is not). This date is based on 30 microsatellite loci on chromosomes 13 and 15, many of which are in close linkage.

In spite of their increasing use in many areas of biology, the mutational process that generates polymorphism at microsatellite loci is poorly understood. Shriver *et al.* (1993) and Valdes *et al.* (1993) used computer simulations to show that mutations increase or decrease allele size by one repeat unit, as in the stepwise mutation model of Ohta and Kimura (1973). However, from empirical data, the mutational process is not exclusively "one-step", suggesting that sometimes mutations change allele size by several repeat units (Di Rienzo *et al.* 1994). Current models assume no constraints on allele size, and as a consequence they predict that the difference in the average number of repeats in reproductively isolated populations tends to increase without bound at a rate proportional to their time of separation (Goldstein *et al.* 1995*b*; Slatkin 1995). However, it appears likely that there are constraints that restrict microsatellite variation to bounded intervals

(Bowcock *et al.* 1994; Deka *et al.* 1994; Garza *et al.* 1995; Rubinsztein *et al.* 1995*a*; Nauta and Weissing 1996; Feldman *et al.* 1997; Zhivotovsky *et al.* 1997). One such constraint may be mutational bias such that large alleles mutate preferentially to smaller alleles, and vice versa for small alleles. This aspect of microsatellite evolution is potentially troublesome for inferring molecular divergence dates.

Since the underlying molecular evolutionary processes of microsatellite variation are still being discovered along with estimates for their basic parameters (*e.g.*, mutation rates) (Rubinsztein *et al.* 1995*a*, 1995*b*; Crouau-Roy *et al.* 1996; Chakraborty *et al.* 1997), current conclusions based on microsatellite data should be viewed tentatively.

1.4.4 Nuclear DNA Sequence Data

Not withstanding the advantages of RFLP and tandem repeat loci over blood groups and proteins, it should be remembered that these newer classes of polymorphisms also require some extrapolation to measure genetic variation at the nucleotide level, since not all nucleotide changes are directly measured by these techniques. The ultimate method of measuring genetic variation is the direct study of DNA sequences.

Rapacz *et al.* (1991) used immunological and DNA sequence analysis to examined the apolipoprotein B locus in human populations. Using chimpanzee and gorilla outgroups to identify the ancestral human haplotype, they concluded an African origin because the frequency of this ancestral haplotype progressively declines as one goes from Africa to the Far East. However, Breguet *et al.* (1990) performed a detailed analysis of the haplotype frequencies and concluded that Caucasoid populations (located from North Africa to India) were closest to the ancestral genetic stock and that worldwide genetic differentiation at this locus is best explained by westward and eastward gene flow from this geographical area and not by an African origin. Xiong *et al.* (1991) studied two apolipoprotein deficiency alleles, one from Japan and one from Venezuela. The data suggest that the two variants arose more than 500,000 years ago. Their persistence since then at a very low frequency shows that there was no population bottleneck over that period, although this has been disputed (Takahata 1993*a*; Rogers and Jorde 1995).

DNA sequence data from the γ^1-γ^2 globin region (Bailey *et al.* 1992) show that the degree of divergence between human alleles is about a fifth that between human and chimpanzee alleles. Assuming a human-chimpanzee divergence time of 6 Ma, this implies that human γ-globin sequences diverged 1.2 Ma ago. However, in this comparison, one of the human sequences is a composite sequence from two individuals. Furthermore, selection may well be acting on γ-globin genes so that alleles with heterozygote advantage could be maintained longer than under neutrality. The noncoding ψη-globin pseudogene is presumably selectively neutral, and human sequences diverged approximately 1.3 Ma ago (Bailey *et al.* 1992). However, as in the γ-globin case, the two available human sequences are composites of two alleles. A region 5′ to the δ-globin gene provides an even older molecular divergence date of 3 Ma between human sequences (Maeda *et al.* 1983). Takahata (1993*a*) argues that the probability of two neutral alleles persisting over 3 Ma is very low, assuming an N_e of 10^4 and no gene conversion or unequal crossing-over occurred. This suggests that either the polymorphism is not neutral or that N_e is $>10^4$ (see Section 1.4.5 below). Although these data are suggestive of a relatively old molecular divergence date, complete sequences from single human alleles are needed to explore this further, as well as errors on the inferred molecular divergence date (Ruvolo 1996). Recently, a 3-kb region encompassing the β-globin gene has been analysed for allelic sequence polymorphism in nine populations from Africa, Asia and Europe (Harding *et al.* 1997). Concordant with other studies, the ancestral sequence for the total sample was found only in Africa. However, diversity differences between Africa and Asia are better explained by greater African effective population size than by greater time depth in Africa. This may due to population-size reduction in Asia rather than population expansion in Africa. They conclude that there has been extensive worldwide late Pleistocene gene flow and that modern humans have both African and Asian ancestry dating to >200,000 years ago.

1.4.5 Major Histocompatibility Complex Loci

The major histocompatibility complex (MHC) consists of several very polymorphic loci which have been used to evaluate competing models of modern human origins. For example, polymorphism at the human *DRB1* locus has an estimated coalescence time of

60 Ma (Ayala 1995). The *DQB1* locus shows a similar pattern of polymorphism (Ayala and Escalante 1996). The nature of the polymorphism indicates that the long-term effective population size of humans was as large as 10^5 (*e.g.*, Takahata 1991, 1993*a*; Ayala *et al.* 1994; Ayala 1995; Ayala and Escalante 1996). Many human alleles have their closest relatives in other primate species (Gyllensten *et al.* 1991), further suggesting that the human lineage has never experienced a severe bottleneck since it first diverged from African apes. However, a high rate of mutation and convergent evolution (Titus-Trachtenberg *et al.* 1994; Gyllensten *et al.* 1996) may partly explain the similarities between human and primate alleles.

1.4.6 *Alu* Insertions

Alu insertions are short (300 bp) primate-specific genetic elements that originated through the process of retrotransposition from a small number of source genes (see review by Deininger and Batzer 1993). Insertion of *Alu* elements is a stable evolutionary event, and no molecular process has been characterised which specifically removes them from the genome (see Hammer 1994, for references and discussion). They have been observed to be removed "nonspecifically" from the genome, in which case, pieces of the original *Alu* element are left behind, or some flanking sequence is removed along with the *Alu* element (Ruvolo 1996).

The *Alu* family is comprised of over 500,000 members which are divided into discrete subfamilies based upon shared mutations (Slagel *et al.* 1987; Willard *et al.* 1987). Several *Alu* family members are so recent in origin that they are polymorphic in the human gene pool (*e.g.*, Batzer and Deininger 1991; Batzer *et al.* 1991, 1994, 1996; Perna *et al.* 1992; Hammer 1994). These recent *Alu* insertions provide a novel set of highly informative nuclear DNA markers for the study of human population genetics since they represent relatively stable polymorphisms that are identical by descent, and the ancestral state of the polymorphism is known (Batzer *et al.* 1994).

Batzer *et al.* (1994) screened 664 individuals from 16 worldwide populations for four polymorphic *Alu* insertions. A maximum-likelihood tree of population relationships

showed four major groupings consisting of Africa, Europe, Asia/Americas, and Australia/New Guinea, which is concordant with similar trees based on classical markers (Cavalli-Sforza *et al.* 1988; Nei and Roychoudhury 1993), nuclear RFLP loci (Bowcock *et al.* 1991*b*; Mountain and Cavalli-Sforza 1994), microsatellite loci (Bowcock *et al.* 1994), and mitochondrial DNA (Merriwether *et al.* 1991). A particularly useful feature of the polymorphic *Alu* insertions is that the ancestral state is known to be the absence of the insertion (which is supported by the absence of these *Alu* insertions at orthologous positions in nonhuman primate genomes). The root of the tree can then be obtained by including a hypothetical ancestral population in the analysis, with the frequency of each insertion set to zero. The results of such analyses indicate that the most probable placement of the ancestral population is within the African branch, suggesting an African origin for these four polymorphic *Alu* insertions. Batzer *et al.* (1996) studied the distribution of six polymorphic *Alu* insertions in 563 individuals from 14 human populations, and concluded (as above) that these insertions probably have an African origin. In a similar study, Stoneking *et al.* (1997) studied the distribution of eight polymorphic *Alu* insertions in 1500 individuals from 34 worldwide populations. In accordance with other genetic data, they found that African populations exhibit the most between-population diversity, and the population tree is rooted in Africa. Furthermore, this study provides evidence that greater African diversity is the result of a larger long-term effective population size in Africa (Relethford and Harpending 1994, 1995; Relethford 1995; Harpending *et al.* 1996). While these data support an African origin, again they do not distinguish between the alternative models of human evolution.

Polymorphic *Alu* insertions represent a unique source of genetic variation for human population genetics. When combined with other genetic information in the region around an *Alu* insertion site, they can provide important information about times of human population divergence (Ruvolo 1996). Recently, a human *Alu* deletion polymorphism and a short tandem repeat polymorphism (STRP) at the CD4 locus on chromosome 12 were analysed in more than 1,600 individuals from 42 worldwide populations (Tishkoff *et al.* 1996). The slow-evolving *Alu* deletion polymorphism provides a timing mechanism, while the faster-evolving STRP differentiates among populations. Sub-Saharan African populations had more haplotypes and exhibited more variability in haplotype frequencies

than the Northeast African or non-African populations. The *Alu* deletion was nearly always associated with a single STRP allele in non-African and Northeast African populations but was associated with a wide range of STRP alleles in the sub-Saharan African populations. Making a number of assumptions, including a constant mutation rate, Tishkoff *et al.* (1996) concluded that the pattern of haplotype variation and the degree of linkage disequilibrium support a recent common African origin for all non-African human populations, perhaps 100,000 years ago. However, this system represents essentially a single locus which has been criticised for its lack of statistical reliability (Pritchard and Feldman 1996). Furthermore, Hammer *et al.* (1997) suggest that the data may have traced a more recent migration, consistent with the multiple African dispersal hypothesis proposed by Lahr and Foley (1995).

1.4.7 The Y Chromosome

The Y chromosome is seen as the male counterpart of mtDNA. Although, a portion of the Y chromosome, known as the pseudoautosomal region, is homologous to and can recombine with the X chromosome, the rest of the Y chromosome appears to be haploid and paternally inherited. It should therefore be possible to trace paternal lineages in human evolution by analysing Y chromosomes (see reviews by Jobling and Tyler-Smith 1995; Hammer and Zegura 1996; Mitchell and Hammer 1996). However, two problems have delayed progress in using the Y chromosome in evolutionary studies. First, earlier RFLP searches (Jakubiczka *et al.* 1989; Malaspina *et al.* 1990; Spurdle and Jenkins 1992) and recent searches using sequencing (Seielstad *et al.* 1994; Dorit *et al.* 1995; Hammer 1995; Whitfield *et al.* 1995), denaturing high-performance liquid chromatography (DHPLC) (Underhill *et al.* 1996), and single-stranded conformation polymorphism (SSCP) (Zerjal *et al.* 1997) have revealed an apparent lack of variability on the Y chromosome. Second, until recently, most available Y-specific polymorphisms have been laborious to type and of a complex molecular nature, so that the collection of data and the design of models with which to interpret them have proved difficult (Lucotte *et al.* 1989; Oakey and Tyler-Smith 1990; Torroni *et al.* 1990; Spurdle and Jenkins 1991; Mathias *et al.* 1994). Thus, Y chromosome markers that are easy to type and of a clear-cut molecular nature would be desirable for population studies. Fortunately such markers

are now becoming available. Linares *et al.* (1996) typed five polymorphic markers on the Y chromosome (mostly microsatellites) in 121 individuals from 13 worldwide populations and found that the greatest genetic distance was between African and non-African populations. Significant geographic clustering of haplotypes was also observed, suggesting that microsatellite-based Y-haplotype trees might have more structure than those seen with mtDNA. Using DHPLC methodology, Underhill *et al.* (1997) identified 19 previously unreported Y biallelic polymorphisms in 718 diverse chromosome. A more detailed Y chromosome phylogeny is emerging which appears to contain a non-African ancestral lineage.

A simple polymorphism resulting from the recent insertion of an *Alu* element on the long arm of the Y chromosome has proved to be useful in human population studies (Persichetti *et al.* 1992; Hammer 1994; Spurdle *et al.* 1994). This element, referred to as the Y *Alu* polymorphic (YAP) element, is present at a specific site (locus *DYS287*) in some individuals and absent in others. The frequency of the YAP insertion has been determined in 340 individuals from 14 worldwide populations (Hammer 1994), and the results combined with other surveys (Persichetti *et al.* 1992; Spurdle *et al.* 1994). Sub-Saharan African populations have the highest frequencies, followed by populations from northern African, Asia, Europe, the New World and Oceania (Hammer 1994). The greatest genetic distance is observed between the African and non-African populations, consistent with the vast majority of molecular studies mentioned so far. Thus, it was proposed that the YAP element originally inserted on an African Y chromosome and subsequently spread to other continents (Hammer 1994).

Depending on the exact timing of the insertion event, the distribution pattern may help to distinguish among hypotheses on the origin of modern humans. Hammer (1995) sequenced a 2.6 kb region, encompassing the YAP insertion, from 16 geographically diverse humans and four chimpanzees. The data strongly suggest that the YAP element inserted on the human Y chromosome a single time after the divergence of humans and chimpanzees. The estimated time back to the ancestral Y chromosome was 188,000 years with a 95% confidence interval of 51,000 to 411,000 years. In a further study of human Y chromosome variation, Hammer *et al.* (1997) examined seven polymorphic

markers (including the YAP element) in 1,500 males from 60 worldwide populations. Population trees based on Y chromosome haplotypes show that African populations are clearly differentiated from the rest of the world. Furthermore, African populations exhibit the highest haplotype diversity, and haplotype variation outside of Africa is a subset of the variation existing within Africa. These findings parallel results from a large variety of other loci which have been interpreted to support an African origin model. However, Hammer *et al.* (1997) suggest that patterns of haplotype diversity are compatible with a variety of hypotheses, including multiple human migrations (Lahr and Foley 1995). The significantly higher level of diversity associated with one haplotype (haplotype 3) in Asia is indicative of a possible Asian origin. Since this haplotype is thought to be ancestral to two of the most common ones found in Africa, they conclude that a major component of the African paternal gene pool is derived from Asia *contra* Hammer (1994). A single nucleotide polymorphism in the SRY region also appears to have originated in Asia (Altheide and Hammer 1997). Similar patterns of variation at the β-globin locus further support the hypothesis of an ancient migration of human populations from Asia to Africa (Harding *et al.* 1997).

Additional Y-specific polymorphisms should facilitate estimates of the coalescence time for all human Y chromosomes. Recent studies have attempted to estimate the coalescence time for Y-haplotypes based on sequence data. Dorit *et al.* (1995) sequenced a 729 bp intron of the *ZFY* gene and found no variation in a worldwide sample of 38 males. From this they estimated an expected coalescence time of 270,000 years, with a 95% confidence interval of 0 to 800,000 years. Whitfield *et al.* (1995) sequenced >15 kb from the non-recombining portion of five human Y chromosomes and found only three polymorphic sites. The Y coalescence time was estimated to be 37,000 to 49,000 years. A recent coalescence time could be due to a recent origin for modern humans, recent male population bottlenecks, extensive male migration, historically small effective male population sizes, or a recent selective sweep. The reduced level of Y chromosome polymorphism in humans and other primate species relative to the nucleotide sequence divergence between these species is consistent with selection acting on the *ZFY* gene or linked loci (Burrows and Ryder 1997). This renders problematical the use of these data in establishing Y chromosome divergence times. Considerable criticisms have been

leveled at attempts to estimate coalescence dates from Y chromosome data (Donnelly *et al.* 1996; Fu and Li 1996; Rogers *et al.* 1996; Weiss and von Haeseler 1996; Tavaré *et al.* 1997). Criticisms have focused principally on varying assumptions about mutation rate, the effective population size, and the dynamics of population growth throughout human history, as well as the limits of the database that is available at present (Mitchell and Hammer 1996; Wolpoff and Caspari 1997).

1.5 DISCORDANCE BETWEEN MITOCHONDRIAL AND NUCLEAR DATA

The above review of the genetic evidence shows that a variety of molecular datasets support an African origin for modern *H. sapiens*. The data have confirmed an African/non-African split and revealed increased genetic diversity in African populations. However, as noted previously, these observations do not distinguish between the alternative models of human evolution. Patterns of mtDNA variation are consistent with a rapid human population expansion (*e.g.*, Di Rienzo and Wilson 1991; Rogers and Harpending 1992; Harpending *et al.* 1993; Sherry *et al.* 1994; Rogers and Jorde 1995), following a period of small population size about 100–200 ka ago (Cann *et al.* 1987*a*; Vigilant *et al.* 1991). This historical portrait has been widely interpreted as support for the replacement model. However, the data are inconclusive as mtDNA is essentially a single genetic locus and may not be selectively neutral (Whittam *et al.* 1986; Excoffier 1990; Merriwether *et al.* 1991; Jorde *et al.* 1995; Nachman *et al.* 1996; Templeton 1996; Wise *et al.* 1998). Moreover, limited nuclear DNA studies reveal somewhat older molecular divergence dates (Maeda *et al.* 1983; Bailey *et al.* 1992; Takahata 1993*a*). It could be that human populations have not, as a whole, recently expanded from an African origin, but that a selectively favoured mtDNA type has. The issue can only be resolved if we have more genes with which to compare the mitochondrial data.

If human mitochondrial genome diversity reflects the species' history of population size and structure, these will be similarly reflected in the pattern of nuclear genome diversity. However, if natural selection is the primary cause of the mtDNA diversity pattern, then other genes may not reveal the same pattern.

1.5.1 Comparison of Mitochondrial and Nuclear Variation Within Humans

The distribution of DNA polymorphism frequencies (frequency spectrum) can be used to ask questions about recent natural selection and nonequilibrium conditions, such as changes in population size (Tajima 1989*a*, 1989*b*; Rogers and Harpending 1992; Fu and Li 1993). Studies of human mtDNA sequences have revealed an excess of polymorphisms with a low-high frequency pattern (*i.e.*, of the two segregating bases at a polymorphic site, one base has a low frequency and the other base a high frequency), and a paucity of intermediate frequency polymorphisms (Johnson *et al.* 1983; Whittam *et al.* 1986; Excoffier 1990; Merriwether *et al.* 1991; Jorde *et al.* 1995; Nachman *et al.* 1996; Wise *et al.* 1998). In contrast, nuclear genes tend to show an excess of intermediate frequency polymorphisms (Hey 1997). These differences could be due to the different populations sampled for the different studies.

To examine mtDNA-nuclear differences more closely, Hey (1997) sequenced a portion of the X-linked pyruvate dehydrogenase E1α subunit (*PDHA1*) locus and a portion of the mitochondrial control region from four sub-Saharan African and four non-African individuals. The two genes revealed a significant difference in the frequency spectrum. *PDHA1* revealed an excess of intermediate frequency polymorphisms, while the control region showed an excess of polymorphisms with the low-high frequency pattern. Differences in the frequency distributions were assessed using Tajima's (1989*a*) *D* test. *D* is proportional to the difference between two commonly used measures of nucleotide variation, π and θ (see Sections 2.6.3 and 2.6.4). π is the average number of pairwise nucleotide differences between sequences (Tajima 1983). A polymorphism of intermediate frequency appears as a difference between many pairs of sequences, and so contributes more to π than does a polymorphism with the low-high frequency pattern. In contrast, θ is a function only of the number of DNA sequences (n) and the number of polymorphic sites (S), and it does not depend on the frequency of polymorphisms (Hudson 1990). For the mtDNA data, the excess of low frequency polymorphisms results in a π value that is less than θ, and *D* is negative (see Chapter 4; Jorde *et al.* 1995; Nachman *et al.* 1996; Hey 1997; Wise *et al.* 1998). In the case of nuclear genes there is an excess of intermediate frequency polymorphisms, and *D* is positive (Hammer

1995; Hey 1997). The discrepancy led to the conclusion that mitochondrial variation has been shaped by natural selection. The positive *D* values from nuclear genes suggest that human populations were not subject to recent population bottlenecks (Hey 1997).

1.5.2 Comparison of Mitochondrial and Nuclear Variation Between Humans and Chimpanzees

The assumption of selective neutrality of the human mitochondrial genome can also be tested by comparing the relative levels of diversity in the mitochondrial genomes of humans and other related primate species. If neutrality is assumed and humans are found to have low levels of mitochondrial genome diversity relative to related species, they should also have relatively low levels of nuclear genome diversity. On the other hand, if humans have reduced mitochondrial genome diversity as a result of natural selection, this would not result in any substantial reduction of nuclear genome diversity. In this case, a reduced level of mitochondrial genome diversity in humans relative to related species would not be matched by a similarly lower level of nuclear genome diversity.

An examination of this hypothesis reveals that humans have a substantially lower ratio of mitochondrial to nuclear genome diversity than chimpanzees and other non-human primates (see Chapter 3; Wise *et al.* 1997). This discrepancy is difficult to reconcile with any neutral model of human mitochondrial genome evolution.

1.6 THE NEUTRAL THEORY

The neutral theory of molecular evolution asserts that most mutations are deleterious and are quickly removed from the population thereby contributing little, if anything, to the levels of polymorphism detected within a species (Kimura 1983). Genetic variation within a species is largely due to random genetic drift of mutations that are selectively equivalent (*i.e.*, neutral). The neutral theory provides a useful null hypothesis against which to assess the significance of the different forces influencing genetic variation within and between species because it makes simple, testable predictions.

Hudson *et al.* (1987) devised a test which compares levels of interspecific divergence in two or more regions of DNA with levels of intraspecific polymorphism in the same regions from at least one species. A modification of this test was pioneered by McDonald and Kreitman (1991). They proposed a simple, and very elegant, approach based on the neutral theory in which the ratio of amino acid replacement (nonsynonymous) to silent (synonymous) polymorphisms within species should be the same as the ratio of non-synonymous to synonymous fixed differences between species. This lends itself to a 2×2 contingency test of homogeneity. However, as pointed out by Templeton (1996), there is no inherent limit in the contingency approach to just two mutational categories. There are few assumptions required because forces that affect polymorphism and divergence are expected to affect equally both nonsynonymous and synonymous sites, but it is assumed that there are no multiple substitutions (McDonald and Kreitman 1991), and, though not explicitly stated, that mutations are independent.

A number of recent studies have used the McDonald and Kreitman approach to test the hypothesis that mtDNA variation is neutral, and this forms part of the present study (see Chapter 4; Wise *et al.* 1998). Nachman *et al.* (1994) sequenced the NADH dehydrogenase subunit 3 (*ND3*) gene from 56 wild *Mus domesticus* and compared polymorphism within species to divergence between *M. domesticus*, *M. musculus* and *M. spretus*. They found that the ratio of nonsynonymous to synonymous nucleotide differences was significantly greater within species than between species, indicating incompatibility with a strictly neutral model of mtDNA evolution. In *Drosophila*, non-neutral patterns have been documented for NADH dehydrogenase subunit 5 (*ND5*) (Rand *et al.* 1994) and cytochrome *b* (*cyt b*) (Ballard and Kreitman 1994). Again, common to these studies was the finding of higher ratios of nonsynonymous to synonymous nucleotide differences within species than between species either for all or part of the genes in question. Similar patterns have also been observed for the *ND3* (Nachman *et al.* 1996) and cytochrome *c* oxidase subunit II (*COII*) (Templeton 1996) genes in humans and chimpanzees. For discussion of other studies that have used the McDonald and Kreitman approach, see the brief review by Brookfield and Sharp (1994).

1.7 AIMS OF THE PROJECT

A basic aim of population genetics is to understand the forces that shape and maintain genetic variability in nature. This is important not only for understanding how evolution proceeds, but also for choosing appropriate genetic markers for population studies. Mitochondrial DNA is used extensively as a marker in population and evolutionary studies, and is generally assumed to evolve according to a neutral model of molecular evolution. This assumption is important since directional selection at any locus in the mitochondrial genome would, in the absence of recombination, reduce variability in the entire genome (Maynard-Smith and Haigh 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992; Braverman *et al.* 1995). Background selection against slightly deleterious mutations may also play a role in the reduction of linked neutral polymorphisms in regions of low recombination (Charlesworth *et al.* 1993, 1995; Charlesworth 1994).

Non-neutrality of mtDNA has wide-ranging evolutionary implications, and provides an interesting alternative hypothesis to be tested. As mentioned previously this can be done in several different ways (see Sections 1.5 and 1.6). In Chapter 3 the assumption of selective neutrality of the human mitochondrial genome is examined by comparing mitochondrial and nuclear genome diversity in humans and chimpanzees (see Section 1.5.2). In Chapter 4 contingency tests of the nonsynonymous to synonymous ratio within and between species (see Section 1.6), and Tajima's (1989*a*) and Fu and Li's (1993) tests are used to investigate departures from neutrality within the mitochondrial genomes of humans and chimpanzees, in particular the *ND2* and *cyt b* genes. Finally, chapter 5 summarises the findings of this thesis, and suggests areas that deserve further exploration and research.

2.1 SAMPLES

2.1.1 Chimpanzee Samples

Common chimpanzee (*Pan troglodytes*) blood samples were obtained from animals held under long-term observation in one of several primate colonies at the Laboratory of Slow, Latent and Temperate Virus Infection of the National Institutes of Health (NIH). They were supplied by Drs. D. C. Gajdusek and C. J. Gibbs Jr. All individuals analysed in this thesis were drawn from a larger sample of 102 individuals, the majority of which were wild-caught (Board *et al.* 1981). Three major subspecies of *P. troglodytes* are currently recognised: *P. t. troglodytes*, *P. t. schweinfurthii*, and *P. t. verus*. In the wild, these subspecies are geographically isolated yet very similar morphological. The geographic origin of most captive chimpanzees in the United States, including those in our sample, is unknown. Previous analysis of mitochondrial control region sequences (see Section 3.4.1; Wise *et al.* 1997) indicates, however, that almost all of the individuals included in the present study are from the west African subspecies *P. t. verus*. Samples were stored as blood clots at –70°C until required.

2.1.2 Human Samples

Human blood samples were obtained from a variety of sources. Samples used in the analysis of tandem repeat loci (see Chapter 3) were obtained from Caucasoid laboratory workers in the John Curtin School of Medical Research (JCSMR). For analysis of the mitochondrial *ND2* and *cyt b* genes (see Chapter 4) samples were obtained from indigenous populations of Africa (Bantu speaking people from South Africa), Asia (Cantonese from Hong Kong), Australia (Aboriginal Australians from the Kimberley region of Western Australia), and Europe (Anglo-Celts from JCSMR). The Asian and European samples were collected as whole blood. The buffy coats were removed after centrifugation and stored at –70°C until required. The Aboriginal Australian and African samples were stored as haemolysates at –70°C until required.

2.2 DNA EXTRACTION

2.2.1 Chimpanzee Samples

The chimpanzee DNA samples analysed in this thesis were prepared in previous studies using a salt extraction method (Miller *et al.* 1988). A number of these samples yielded poor quality DNA. For the work of this thesis, higher quality DNA was prepared using a modification of the method described by Grunebaum *et al.* (1984). Blood clots (1–2ml) were made up to 4ml with $T_{20}E_5$ (20mM Tris-HCl, 5mM EDTA pH 8.0) and incubated at 37°C overnight in the presence of 200µl 10% ultrapure SDS and 50µl Proteinase K (20mg/ml). The DNA was then purified by successive extractions with equal volumes of phenol (saturated with $T_{10}E_1$ buffer: 10mM Tris-HCl, 1mM EDTA pH 8.0), equal volumes of phenol/chloroform/isoamyl alcohol (25:24:1), and equal volumes of chloroform/isoamyl alcohol (24:1) (modified after Maniatis *et al.* 1982). Due to the high proportion of red blood cells present, some samples required up to six initial phenol extractions. DNA was precipitated by the addition of 0.1 volumes of 5M NaCl and 2.5 volumes of cold absolute ethanol and incubation at –20°C overnight. Following 10 minutes centrifugation, the DNA pellet was rinsed in 70% cold ethanol and re-centrifuged. The resulting pellet was air dried and resuspended in an appropriate volume (100–300µl) of ddH_2O.

2.2.2 Human Samples

The Asian and European DNA samples were prepared in previous studies from buffy coats using the method of Grunebaum *et al.* (1984) (see Section 2.2.1). Genomic DNA from the Aboriginal Australian and African samples was extracted using the ABI 340A Nucleic Acid Extractor (Applied Biosystems, Foster City, USA). A modification of the standard buffy coat protocol for 30ml vessels was used. This involved digestion with Proteinase K for up to four hours (instead of one hour) and three to four initial phenol/water/chloroform extractions (instead of two). All DNA extracted on the automated extractor was resuspended from precipitate filters in an original volume of 300µl TE with overnight incubation at 55°C.

DNA concentrations were determined by measuring the optical density of solutions at 260nm (1 OD_{260} unit = 50µg/ml for double stranded DNA) (Maniatis *et al.* 1982) using a CARY-1 UV spectrophotometer. DNA samples for PCR were diluted with ddH_2O to a final concentration of 25ng/µl and stored at 4°C.

2.3 POLYMERASE CHAIN REACTION (PCR)

The polymerase chain reaction (Saiki *et al.* 1988) was used extensively during the course of this study to selectively amplify the mitochondrial NADH dehydrogenase subunit 2 (*ND2*) gene, the cytochrome *b* (*cyt b*) gene and the first hypervariable region (*segment I*) of the control region. Tandem repeat loci were also amplified by PCR as required for further analysis (see Section 2.5).

2.3.1 Primers

Oligonucleotide primers for PCR and sequencing were synthesised by the Biomolecular Resource Facility, JCSMR, Australian National University, using a 3-column ABI 380B Oligonucleotide Synthesiser. Primers were supplied in ammonia solution. These were aliquoted into 200µl fractions, vacuum centrifuged until dry, rinsed in 100µl ddH_2O, redried and resuspended in 75µl ddH_2O. The concentration of primer solutions was then measured (1 OD_{260} unit = 33µg/ml for single stranded DNA) (Maniatis *et al.* 1982) and the solutions diluted with ddH_2O to 10µM for PCR reactions and 3.2µM for dye terminator sequencing reactions.

2.3.2 Amplification of Mitochondrial DNA

Oligonucleotide primers used for PCR amplification of the mitochondrial control region, the *ND2* gene, and the *cyt b* gene are given in Table 2.1. All reactions were carried out in 50µl volumes. The protocols recommended by the suppliers of *Taq* polymerase (Promega or Advanced Biotechnologies) were followed with some modifications in order to achieve optimal reaction conditions. Initial reaction conditions were: denaturation at 93°C or 96°C, and extension at 73°C or 74°C for a time period determined by the

length of the product and the type of thermal cycler used. The annealing temperature was calculated for each primer (Suggs *et al.* 1981), and the lower temperature was used. Subsequent reaction conditions were modified as necessary to optimise the yield of specific product, usually by altering the annealing temperature or by varying amounts of *Taq*, primer and/or DNA. Thirty to thirty-five cycles were performed for all amplifications. To confirm successful amplification 5µl of PCR product was visualised on a 2% agarose gel (see Section 2.3.3). The remainder of the product was stored at 4°C. Negative controls containing all reaction components except template DNA were included in each PCR run to monitor for contamination. PCR was carried out in an Innovonics Gene Machine (Bartelt Instruments, Victoria, Australia) or during the latter part of the study, a FTS-960 thermal cycler (Corbett Research, Sydney, Australia).

2.3.2.1 Amplification of the control region

Amplification of the first hypervariable region (*segment I*) of the mitochondrial control region was conducted in two stages. The first-stage PCR produced a 1,309 base pair (bp) product encompassing the entire control region. Amplification primers were L15926 and H629 (Chin 1991) (Table 2.1). Amplification was carried out in 50µl reactions containing 100ng genomic DNA, 0.2mM of each deoxynucleotide triphosphate (dNTP), 1.5mM $MgCl_2$, 0.2µM of each primer and 2 units of *Taq* polymerase in reaction buffer supplied by the manufacturer (Promega, Madison, USA). Each reaction was overlaid with mineral oil to prevent evaporation, and PCR was carried out in an Innovonics Gene Machine. After 3 minutes initial denaturation at 96°C, samples were processed through 35 cycles of 75 seconds at 96°C, 75 seconds at 46°C and 105 seconds at 74°C. Products were precipitated by 165µl of cold absolute ethanol and 15µl 3M sodium acetate (pH 5.3) at –20°C for one hour. Following 30 minutes centrifugation, the DNA pellet was rinsed in 70% cold ethanol and re-centrifuged. The resulting pellet was vacuum dried and resuspended in 40µl ddH_2O.

A 1µl aliquot of purified product was used as the template for subsequent nested PCR reactions, with primers based on L15997 (Ward *et al.* 1991) and H16401 (Vigilant *et al.* 1989) (Table 2.1). Concentrations of reagents were as above, with the exception of *Taq* polymerase (0.5 unit). All reactions were subjected to 3 minutes initial denaturation at

96°C prior to 35 cycles of 1 minute at 96°C, 1 minute at 58°C and 1 minute at 74°C. For each individual two separate PCR reactions were performed with the primers (L′ *biotinylated* and H′ *M13 reverse*) and (L′ *–21M13* and H′ *biotinylated*).

2.3.2.2 Amplification of ND2

The mitochondrial NADH dehydrogenase subunit 2 (*ND2*) gene was amplified in three overlapping fragments (I to III). Each fragment was amplified separately in a 50µl reaction containing 100ng genomic DNA, 0.2mM of each dNTP, 1.5mM $MgCl_2$, 0.4µM of each primer and 1.25 units of *Taq* polymerase (Promega) in reaction buffer supplied by the manufacturer. Reactions were overlaid with mineral oil, and PCR was carried out in an Innovonics Gene Machine. Cycling conditions for fragments I and III were 3 minutes initial denaturation at 96°C, and 30 cycles of 1 minute at 96°C, 1 minute at 56°C and 1 minute at 73°C. Fragment II required a lower annealing temperature of 52°C. For each fragment two separate PCR reactions were performed with the primers (L′ *biotinylated* and H′ *–21M13*) and (L′ *M13 reverse* and H′ *biotinylated*).

2.3.2.3 Amplification of cyt b

Amplification of the mitochondrial cytochrome *b* (*cyt b*) gene was carried out in 50µl reaction volumes containing 100–200ng genomic DNA, 1× Advanced Biotechnologies Buffer IV (200mM $(NH_4)_2SO_4$, 750mM Tris-HCl pH 9.0, 0.1% (w/v) Tween), 0.2mM of each dNTP, 1.5mM $MgCl_2$, 0.3µM of each primer and 2 units of *Taq* polymerase (Advanced Biotechnologies). PCR was carried out in a Corbett FTS-960 thermal cycler using the following parameters with a thermal ramping time of 1°C/3 seconds:

[93°C × 3′] × 1 cycle
[93°C × 15″, 54°C × 1′, 73°C × 2′] × 35 cycles
[73°C × 2′] × 1 cycle
hold 4°C.

TABLE 2.1 Primers used for PCR amplification of the mitochondrial control region, the *ND2* gene, and the *cyt b* gene.

Region	Primer	Length (bp)	Sequence (5′ to 3′)
segment I	L15926	19	TACACCAGTCTTGTAAACC
	H629	19	TGTTTATGGGGTGATGTGA
	L15997[a]	20	CACCATTAGCACCCAAAGCT
	H16401[a]	20	TGATTTCACGGAGGATGGTG
ND2			
Fragment I	L4448[a]	19	CCCATACCCCGAAAATGTT
	H4784[a]	19	AAAGGGGGCTATTCCTAGT
Fragment II	L4753[a]	19	CTACCAATCAATACTCATC
	H5166[a]	19	ATGTTAGCTTGTTTCAGGT
Fragment III	L5162[a]	19	CACGACCCTACTACTATCT
	H5529[a]	19	TTGAAGGCTCTTGGTCTGT
cyt b	L14692	22	CATTATTCTCGCACGGACTACA
	H15946	22	CTTTTTCTCTGATTTGTCCTTG

NOTE.—The L or H in the primer name refers to the light or heavy strand, respectively, while the number identifies the base at the 3′ end according to the numbering system of Anderson *et al.* (1981).

[a] Primers with biotin or universal M13 sequence extensions as described in Sections 2.3.2.1 and 2.3.2.2.

2.3.3 Agarose Gel Electrophoresis

PCR products were electrophoresed through 100 × 65 × 5mm agarose gels in TAE buffer (0.04M Tris-Acetate, 0.001M EDTA, pH 8.0) using a Miniphor Submarine Electrophoresis Unit (LKB 2013, Hoefer, San Francisco, USA). Usually 2.0% (w/v) agarose gels were run, although in some cases lower percentage gels (1.0% or 1.5%) were used

in order to separate larger DNA fragments. Prior to electrophoresis PCR product was mixed with Ficoll loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 15% Ficoll 400 in ddH_2O) in a 2:1 ratio (Maniatis *et al.* 1982). Electrophoresis was carried out at 100V for 30–60 minutes, following which the gel was stained in 4µg/ml ethidium bromide for 15–30 minutes, visualised and photographed under UV light (254nm).

2.3.4 Preparation of DNA for Sequencing

2.3.4.1 Single stranded

PCR products required for use as a template in dye primer cycle sequencing reactions (see Sections 2.3.2.1 and 2.3.2.2) were purified from unincorporated primers and nucleotides using the Wizard PCR preps DNA purification system (Promega, Madison, USA) according to the manufacturer's instructions. Single stranded template was then prepared using Dynabeads M-280 Streptavidin (Dynal, Oslo, Norway) according to the manufacturer's protocol with the following modifications. The double stranded product was denatured in 40µl of 0.1M NaOH (instead of 8µl) for 10 minutes. The supernatant containing the non-biotinylated strand was discarded, and the Dynabeads containing the immobilised biotinylated strand were resuspended in 13µl ddH_2O for the sequencing reaction.

2.3.4.2 Double stranded

PCR products required for use as a template in dye terminator cycle sequencing reactions (see Section 2.3.2.3) were purified by passage through a Microcon-100 micro-concentrator (Amicon Division, W.R. Grace, USA) using a modification of the manufacturer's protocol. Briefly, each unit was rinsed with 500µl ddH_2O and then filled with PCR product diluted to 500µl with ddH_2O. This was spun at 3000 rpm for 10 minutes, followed by 3000 rpm for 5 minutes with an additional 200µl ddH_2O. The unit was then inverted and spun for 3 minutes at 1000 rpm to collect the retentate of 40µl.

An alternative procedure was followed in instances where PCR amplification of the desired fragment was accompanied by non-specific amplification. PCR products (20µl) were electrophoresed through 1.0% Sea Plaque (low melting temperature) agarose gels

containing 10μg ethidium bromide using similar conditions to those previously outlined in Section 2.3.3. The target band was excised from the gel under minimal UV illumination to reduce nicking of the DNA. The DNA was then recovered from the agarose using the GELase system (Epicentre Technologies, Madison, USA) according to a modification of the "fast protocol" given by the manufacturer. The gel slice was completely melted by incubation at 70°C for 15 minutes, and the molten agarose digested at 42°C overnight in the presence of 0.2 units of GELase enzyme. Template DNA was then purified as outlined in the above paragraph.

2.4 AUTOMATED DNA SEQUENCING

Cycle sequencing, that is, fluorescence-based DNA sequencing in conjunction with PCR (Smith *et al.* 1986; Gibbs *et al.* 1989), was used to prepare samples for sequence analysis with the ABI Model 373A or ABI Model 377 Automated DNA Sequencers. One of two protocols was employed for sample preparation.

2.4.1 Cycle Sequencing with Dye Primers

Purified single stranded DNA templates (see Section 2.3.4.1) with appropriate forward or reverse universal M13 sequence extensions were sequenced directly by employing either the –21M13 or M13 reverse Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing kits depending on desired sequencing direction. Volumes of reagents used were according to the manufacturer's instructions. Cycle sequencing reactions were carried out in a Corbett FTS-IS capillary thermal cycler using the following parameters with a thermal ramping time of 1°C/2 seconds:

[94°C × 3′, 52°C × 15″, 72°C × 1′] × 1 cycle
[94°C × 15″, 52°C × 15″, 72°C × 1′] × 19 cycles
[94°C × 15″, 72°C × 1′] × 20 cycles
hold at 25°C.

The M13 reverse dye primer required a slightly lower annealing temperature of 51°C. Following cycle sequencing the extension products were precipitated with 85µl of 95% ethanol and 1.5µl 3M sodium acetate (pH 5.3) at –20°C for at least one hour. After 30 minutes centrifugation the DNA pellet was rinsed twice in 70% cold ethanol and vacuum dried.

2.4.2 Cycle Sequencing with Dye Terminators

Purified double stranded PCR product (see Section 2.3.4.2) was used as a template in the dye terminator reaction (Gibbs *et al.* 1989) employing the Applied Biosystems PRISM Dye Terminator Cycle Sequencing Ready Reaction kit . Amounts of all reagents used were exactly as specified by the suppliers. For *cyt b* three internal sequencing primers, at 300–500 bp intervals, were used in each direction (Table 2.2). Reactions were carried out in a Corbett FTS-IS capillary thermal cycler (30 seconds initial denaturation at 94°C, and 25 cycles of 30 seconds at 94°C, 15 seconds at 50°C, and 4 minutes at 60°C) with a thermal ramping time of 1°C/4 seconds. The extension product was precipitated with 50µl of 95% ethanol and 2.0µl 3M sodium acetate (pH 5.3) at –20°C for 30 minutes, rinsed twice in 70% ethanol and vacuum dried.

"Dye Primer" and "Dye Terminator" cycle sequencing products were separated on 360 × 160 × 0.4mm or 360 × 160 × 0.2mm 6% polyacrylamide gels at the Biomolecular Resource Facility, JCSMR, Australian National University, using the ABI Model 373A or ABI Model 377 Automated DNA Sequencers. Polyacrylamide gel preparation, plate preparation and assembly, sample preparation and loading, electrophoretic conditions, and data analysis were according to the appropriate User's Manual. For each individual both H (heavy) and L (light) strands were sequenced at least once in order to obtain consensus sequences.

TABLE 2.2 Primers used for dye terminator cycle sequencing of the *cyt b* gene.

Primer	Length (bp)	Sequence (5′ to 3′)
L14745	18	TTCAACTACAAGAACACC
L15035	18	TTATCTGCCTCTTCCTAC
L15494	18	TATTCTCACCAGACCTCC
H15896	18	GGTGTATTAGTTTATACT
H15471	18	CTGGTGAGAATAGTGTTA
H14983	18	TTAGCGTGAAGGTAGCGG

2.5 TANDEM REPEAT LOCI

Eight tandem repeat loci were examined in this study, including five tetranucleotide repeats and one 31–34 bp repeat originally characterised in humans, and one tetra- and one dinucleotide repeat that had not previously been characterised in either species. A summary of the genetic loci, primer sequences used for amplification, and the optimal PCR conditions is presented in Table 2.3.

2.5.1 PCR

Amplification was conducted in a 25µl reaction volume overlaid with mineral oil, and performed on an Innovonics Gene Machine. Each reaction contained Promega supplied *Taq* polymerase buffer and 0.2mM of each dNTP while the other components ranged as follows: 1.0–3.0mM $MgCl_2$, 0.1–0.4µM of each primer, 0.5–1.0 units *Taq* polymerase and 50–100ng genomic DNA. Optimal conditions for each primer set were determined experimentally. All reaction mixes were subjected to an initial 3 minutes denaturation prior to 30–35 cycles of amplification. Following amplification 3–5µl of PCR product was run on a 1.5% or 2.0% agarose gel (see Section 2.3.3) to visualise the bands before continuing to polyacrylamide gel analysis.

2.5.2 Polyacrylamide Gel Electrophoresis

Polyacrylamide gels were made up to the appropriate concentration (3.5–10%) with 10 × TBE buffer (1M Tris-borate, 1M boric acid, 0.02M EDTA, pH 8.0) and polymerisation was initiated by the addition of 1/150 volumes of ammonium persulphate and 1/1500 volumes of TEMED (N,N,N′,N′ tetramethylethylene diamine). PCR products were fractionated on 160 × 160 × 1.5mm or 320 × 160 × 1.5mm polyacrylamide gels using Hoefer supplied plates and tank apparatus (Vertical Slab Gel Electrophoresis Unit, SE 600, Hoefer) in TBE buffer at 30–40mA until desired separation was achieved. Immediately before loading, 5–10µl of PCR product was mixed with glycerol loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol) in a 2:1 ratio, and the samples were heated at 65°C for 10 minutes. After electrophoresis the gels were stained in ethidium bromide for 30 minutes and visualised under UV light. Alternatively, gels were stained with silver as outlined in the Qiagen protocol (Qiagen, Studio City, USA) and visualised under normal illumination.

Stock solutions of 40% (w/v) acrylamide (19:1 acrylamide/N,N′-methylene-*bis*-acrylamide) (Biorad, Richmond, USA) were stored at 4°C in the dark. Ammonium persulphate (10%) was made weekly and stored at 4°C.

TABLE 2.3 Summary of the eight tandem repeat loci analysed in this study. Conditions given are designed for use on the Innovonics Gene Machine and show the temperature (°C), and the time (sec).

Locus	Primer 1 (5′ to 3′)	Primer 2 (5′ to 3′)	Conditions		
			Denature °C/sec	Anneal °C/sec	Extend °C/sec
COL2A1[a]	CCAGGTTAAGGTTGACAGCT	GTCATGAACTAGCTCTGGTG	94/60	55/60	70/90
D7S460[b]	AATACCCCAAGGGGTGGTAA	CATTGATGAACAGTTCAAGCA	94/90	50/90	74/90
D8S342[c]	CAGCCTGGGCAATAGAAAGAGAC	CAGTGCTCCCTTCCTTGAAGTTTC	95/45	58/30	73/60
MYCN[d]	GGAGGCTGAAAGCACAGTTG	TGGGCAACAAGAGCAAAACT	96/80	52/60	73/70
RENA4[e]	AGAGTACCTTCCCTCCTCTACTCA	CTCTATGGAGCTGGTAGAACCTGA	95/80	56/60	73/70
THO4[f]	CAGCTGCCCTAGTCAGCAC	GCTTCCGAGTGCAGGTCACA	94/105	53/140	72/140
BG01[g]	ATAGACTGGAGTAAAGGAA	CTTCTACTCTGTGAATGGA	94/60	49/60	73/60
EG01[g]	CAAAGTAGTGGGAAGCTGT	ATGGATAGGATAAGTCCCC	96/60	52/60	72/60

NOTE.—References are as follows: [a] Wu *et al.* (1990); [b] Hudson *et al.* (1992); [c] Lu *et al.* (1993); [d] Fougerousse *et al.* (1992); [e] Edwards *et al.* (1992); [f] Polymeropoulos *et al.* (1991); and [g] This study—previously uncharacterised short tandem repeat loci. BG01 is a dinucleotide repeat $(TG)_n$, which starts at position 1479 of the human and chimpanzee β-globin genomic sequence (Savatier *et al.* 1985). EG01 is a tetranucleotide repeat $(GTAT)_n$, which starts at position 2939 of the human and chimpanzee ψη-globin genomic sequence (Miyamoto *et al.* 1987).

2.6 COMPUTATIONAL ANALYSES

2.6.1 Primer Design

Unpublished primers were designed with the aid of the PCR Primer Selection program (Epicentre Software, Pasadena, USA) or the Oligo 4.04 design program (MedProbe AS, Oslo, Norway).

2.6.2 Sequence Alignment

The SeqEd™ 675 DNA Sequence Editor (ABI) program was used for accurate editing and analysis of 373A and 377 data. Forward- and reverse-complement sequences from the same individuals were aligned and compared for confirmation of results, and to resolve any ambiguities or incorrect base calls. Sequences from published sources were obtained from the DDBJ/EMBL/Genbank International Nucleotide Sequence Database through the Australian National Genomic Information Service (ANGIS). Multiple sequence alignments were performed manually using Genetic Data Environment (GDE) 2.2 (Smith *et al.* 1994).

2.6.3 Measures of Genetic Variation

2.6.3.1 Number of segregating sites

The amount of genetic variation within a species can be estimated from the number of segregating (polymorphic) sites (S) using the relation

$$\theta = S/a_1 \qquad (2.1)$$

where $a_1 = \sum_{i=1}^{n-1} 1/i$ and n is the number of DNA sequences examined (Watterson 1975; Equation 5 in Tajima 1989*a*). The variance of θ can be estimated by

$$V(\theta) = \frac{a_1^2 S + a_2 S^2}{a_1^2 \left(a_1^2 + a_2 \right)} \tag{2.2}$$

where $a_2 = \sum_{i=1}^{n-1} 1/i^2$ (Equation 4 in Tajima 1993).

2.6.3.2 Average number of pairwise nucleotide differences

An alternative measure of genetic variation is the average number of nucleotide differences between all possible pairs of sequences. This is defined by

$$\pi = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \pi_{ij} \Big/ n_c \tag{2.3}$$

where π_{ij} is the number of nucleotide differences between sequences i and j, n is the number of DNA sequences sampled from a population, and $n_c = n(n-1)/2$ is the total number of sequence comparisons (Equation 10.6 in Nei 1987). The total variance of π which incorporates the sampling variance as well as other stochastic factors is given by

$$V(\pi) = \frac{n+1}{3(n-1)} \pi + \frac{2\left(n^2 + n + 3\right)}{9n(n-1)} \pi^2 \tag{2.4}$$

(Equation 30 in Tajima 1983). The stochastic variance is that due to chance factors of gene fixation or loss that occur during the evolutionary process (*i.e.*, genetic drift) while the sampling variance is that due to sampling error at the time of the allelic survey.

It should be noted that as measures of genetic variation, θ and π are only valid when the samples are drawn from a randomly mating population at equilibrium for mutation and genetic drift. Furthermore, both θ and π depend on the length of the DNA sequence (m), and the amount of DNA variation per site can be used instead, which is obtained by dividing θ or π by m.

2.6.3.3 Gene diversity (heterozygosity)

An unbiased estimate of heterozygosity or *gene diversity* for a locus is given by

$$h = 2n(1-A)/(2n-1) \tag{2.5}$$

where n is the number of individuals sampled, and A is either the homozygosity estimated directly by gene counting or calculated as $\sum_{i=1}^{m} x_i^2$ (where x_i is the estimated frequency of the ith allele in the system and m is the number of alleles) (Equation 2 in Nei 1978). The average heterozygosity (H) is estimated by the average of h over all loci studied. The standard error of this estimate can be calculated according to Nei and Roychoudhury (1974*b*).

2.6.4 The Tajima and the Fu and Li Tests of Neutrality

Under neutrality and constant population size, nucleotide variation is determined by the population mutation parameter, which for autosomal regions is $4N_e\mu$ (where N_e is the effective population size and μ is the neutral mutation rate). Since mtDNA is effectively haploid and maternally transmitted, $2N_f\mu$ (where N_f is the effective female population size) is actually more appropriate.There are a number of different ways of estimating this parameter; either from S, the number of segregating (or polymorphic) sites (Watterson 1975), or π, the average number of pairwise nucleotide differences (Nei 1987), or the total number (η) of mutations and the number (η_e) of mutations in the external branches (Fu and Li 1993). These quantities have been used to devise statistical tests of neutrality such as those of Tajima (1989*a*)

$$D = \frac{\pi - S/a_1}{\sqrt{\mathrm{Var}[\pi - S/a_1]}} \tag{2.6}$$

and Fu and Li (1993)

$$D = \frac{\eta - a_1 \eta_e}{\sqrt{\mathrm{Var}[\eta - a_1 \eta_e]}} \tag{2.7}$$

$$F = \frac{\pi - \eta_e}{\sqrt{\mathrm{Var}[\pi - \eta_e]}} \tag{2.8}$$

The expectation and variance of these test statistics under neutrality are approximately zero and one, respectively.

Chapter 3

COMPARATIVE NUCLEAR AND MITOCHONDRIAL GENOME DIVERSITY IN HUMANS AND CHIMPANZEES

3.1 ABSTRACT

The dynamics of human evolution are examined by comparing mitochondrial and nuclear genome diversity between humans and other closely related species, particularly chimpanzees. Restriction mapping and sequencing have shown that humans have substantially lower levels of mitochondrial genome diversity (π) than chimpanzees. In contrast, humans have substantially higher levels of heterozygosity (H) at protein-coding loci, suggesting a higher level of diversity in the nuclear genome. To investigate the discrepancy further, the nucleotide sequence of the first hypervariable region of the mitochondrial control region has been determined for 49 common chimpanzees. The majority of these were from the *Pan troglodytes verus* subspecies, which was under-represented in previous studies. For a total sample of 115 chimpanzees, $\pi = 0.075 \pm 0.037$, compared to 0.020 ± 0.011 for a sample of 1,554 humans. Heterozygosity values at 60 tandem repeat loci have also been estimated in both species. The heterozygosity of human tandem repeat loci is significantly higher than that of chimpanzees. Thus, the higher level of nuclear genome diversity relative to mitochondrial genome diversity in humans is not restricted to protein-coding loci. It seems that humans, not chimpanzees, have an unusual π/H ratio, since the ratio in chimpanzees is similar to that in other catarrhines. This discrepancy in the relative levels of mitochondrial and nuclear genome diversity in the two species cannot be explained by differences in mutation rate. However, it may result from a combination of factors such as a difference in the extent of sex ratio disparity, the greater effect of population subdivision on mitochondrial than on nuclear genome diversity, a difference in the relative levels of male and female migration among subpopulations, diversifying selection acting to increase variation in the nuclear genome, and/or directional selection acting to reduce variation in the mitochondrial genome.

3.2 INTRODUCTION

Comparisons of genetic diversity between closely related species can contribute significantly to a number of issues in population genetics and evolution. Such comparisons have played a role in understanding aspects of genome evolution including the evolutionary dynamics of transposable elements (Dowsett and Young 1982), and the evolutionary patterns and dynamics of short tandem repeat (STR) loci (Rubinsztein *et al.* 1995*a*). They have also been important sources of evidence for the action of natural selection. Balancing selection has been indicated by polymorphism maintained between species (Mayer *et al.* 1992), and discrepancies in levels of diversity within and between species have been interpreted as indicating various forms of natural selection (*e.g.*, McDonald and Kreitman 1991; Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Templeton 1996; Wise *et al.* 1998).

In this context, the increasing wealth of information about different kinds of genetic diversity in humans provides a valuable basis for comparison with closely related species, particularly with chimpanzees. Understandably, the investigation of genetic diversity in chimpanzees has not been as comprehensive as in humans. Most intensively investigated have been electrophoretically detectable protein polymorphisms (King and Wilson 1975; Bruce and Ayala 1979) and mitochondrial genome variation (Ferris *et al.* 1981; Morin *et al.* 1994; Ruvolo *et al.* 1994; Goldberg and Ruvolo 1997). The results have suggested a discrepancy between the relative levels of mitochondrial and nuclear genome diversity in the two species. Chimpanzees appear to have more mitochondrial genome diversity, whereas humans have more nuclear genome diversity.

The ratio of gene diversities in mitochondrial and nuclear genomes, π/H, is approximately $N_f\mu/2N_e\nu$, where N_e is the effective population size, N_f is the number of breeding females, μ is the mutation rate of the mitochondrial genome, and ν is the mutation rate of nuclear genes (Birky *et al.* 1983). This ratio assumes selective neutrality and depends on various factors such as sex ratio, population structure, and mutation rates (Wright 1969). Thus, if there is a discrepancy in the ratio between two species, it is of considerable interest, as it implies a difference in some aspect of their evolutionary dynamics.

In this study, levels of mitochondrial and nuclear genome diversity in humans and chimpanzees are further examined. The number of described chimpanzee mitochondrial control region sequences is extended to include a better representation of the west African subspecies *P. t. verus*, and levels of variation at tandem repeat loci are determined. By investigating this type of nuclear genome variation, which is very different from the protein variation previously studied, it is possible to determine whether the results indicated by the protein variation can be generalised to the nuclear genome.

3.3 METHODS

3.3.1 Hypervariable Region I

DNA sequences from the first hypervariable region of the mitochondrial control region (positions 16,024 to 16,400 in the numbering system of Anderson *et al.* 1981) were obtained from 49 chimpanzees as outlined in Chapter 2. Consensus sequences were produced by aligning forward- and reverse-complement sequences from the same individual in the SeqEd™ 675 DNA Sequence Editor (ABI) program. The 49 chimpanzee control region sequences reported here have been deposited in the DDBJ/EMBL/ Genbank International Nucleotide Sequence Database under accession numbers U84293–U84341. Chimpanzee sequences from published sources were also used in the comparative analyses. Three sequences were obtained from Kocher and Wilson (1991) and 63 were obtained from the DDBJ/EMBL/Genbank International Nucleotide Sequence Database (published in Morin *et al.* 1994) through the Australian National Genomic Information Service (ANGIS). Sequence alignment was performed manually using Genetic Data Environment (GDE) 2.2 (Smith *et al.* 1994). Appendix A shows the alignment of control region sequences from 115 common chimpanzees and one human.

3.3.2 Phylogenetic Trees

The phylogeny of the *P. troglodytes* control region sequences was estimated using the neighbor-joining (NJ) method (Saitou and Nei 1987) as implemented in PHYLIP (Phylogeny Inference Package) 3.5c (Felsenstein 1993). A bootstrapped NJ tree was

constructed using the SEQBOOT program with 1000 replicates. Distance matrices were generated using the program DNADIST with the multiple-hits correction of Kimura (1980). The matrices were used in the NEIGHBOR program with a randomised input order. A consensus tree was created using the CONSENSE program with a *P. paniscus* sequence (Foran *et al.* 1988) as the outgroup.

3.3.3 Pairwise Distances

The numbers of nucleotide differences per site between sequence pairs were estimated using a simple proportional distance (*p*-distance). This was obtained by dividing the number of nucleotide differences between two sequences by the total number of nucleotides compared. To correct for multiple substitutions, between-sequence distances were computed using Kimura's (1980) two-parameter method, with a transition/transversion ratio of nine estimated from the observed sequence data. Distances were also estimated using Tamura and Nei's (1993) method, assuming a Γ distribution with $\alpha = 0.5$ (Wakeley 1993), as implemented in MEGA (Molecular Evolutionary Genetics Analysis) 1.0 (Kumar *et al.* 1993). This distance measure was developed primarily for mitochondrial control region sequence data, where the rate of nucleotide substitution is known to vary from site to site, and there is a strong transition/transversion bias. Alignment gaps and sites with missing information were ignored in the pairwise comparisons. The mean of the uncorrected and corrected pairwise distances or nucleotide diversity (π) was estimated using Equation 2.3 as outlined in Section 2.6.3.2. The standard deviation of this estimate was calculated using Equation 2.4 as outlined in Section 2.6.3.2. The frequency distributions of the pairwise distances were displayed as histograms (mismatch distributions), as proposed by Rogers and Harpending (1992).

3.3.4 Tandem Repeat Loci

The procedures used to analyse eight tandem repeat loci were as outlined in Section 2.5, using the PCR primers and conditions detailed in Table 2.3. Variation in allele length was determined by electrophoretic separation on polyacrylamide gels (Section 2.5.2). Since all loci were autosomal and alleles were codominant, heterozygosity values were esti-

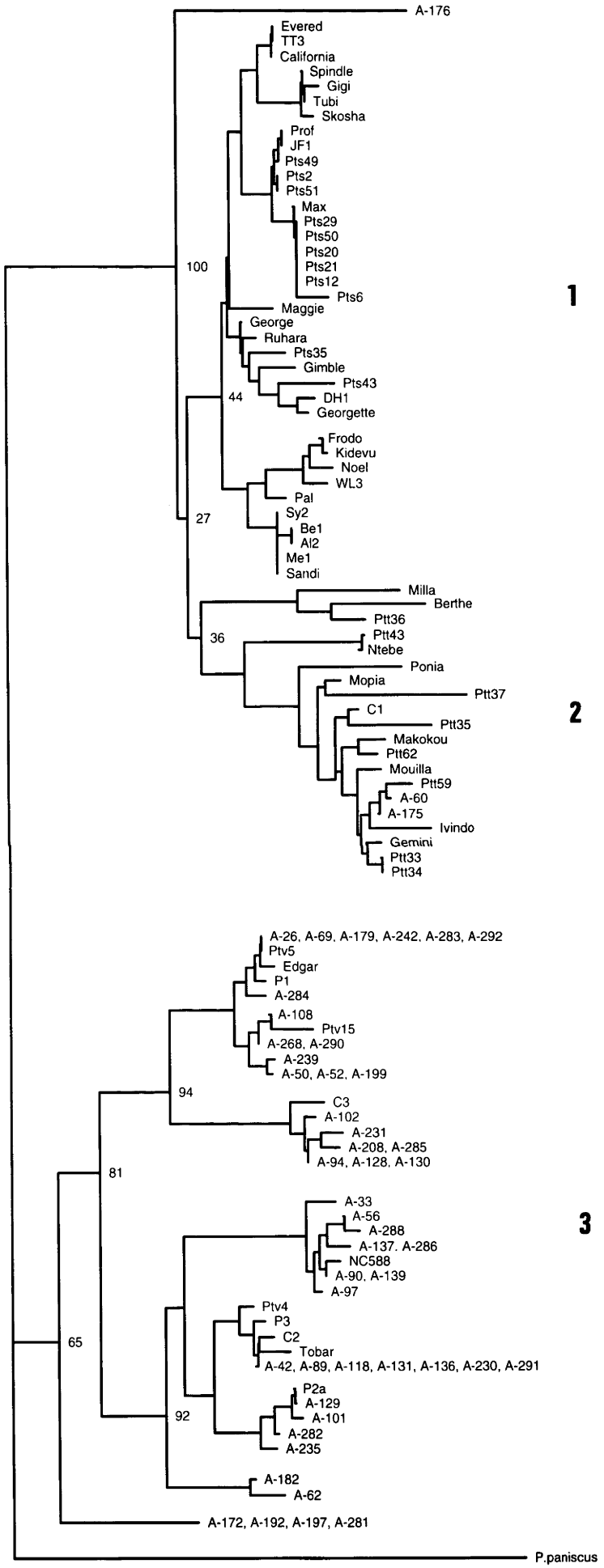
mated directly by gene counting. An additional 60 loci for which variation in humans and chimpanzees had been described elsewhere were included in subsequent analyses. An unbiased estimate of heterozygosity for each locus was calculated using Equation 2.5 as outlined in Section 2.6.3.3. The average heterozygosity (H) is defined as the average of this quantity over all loci examined. The standard error of H was calculated according to Nei and Roychoudhury (1974*b*). Differences in heterozygosity values of human and chimpanzee loci were assessed by the nonparametric Wilcoxon's signed-ranks test (Seigel 1956).

3.4 RESULTS

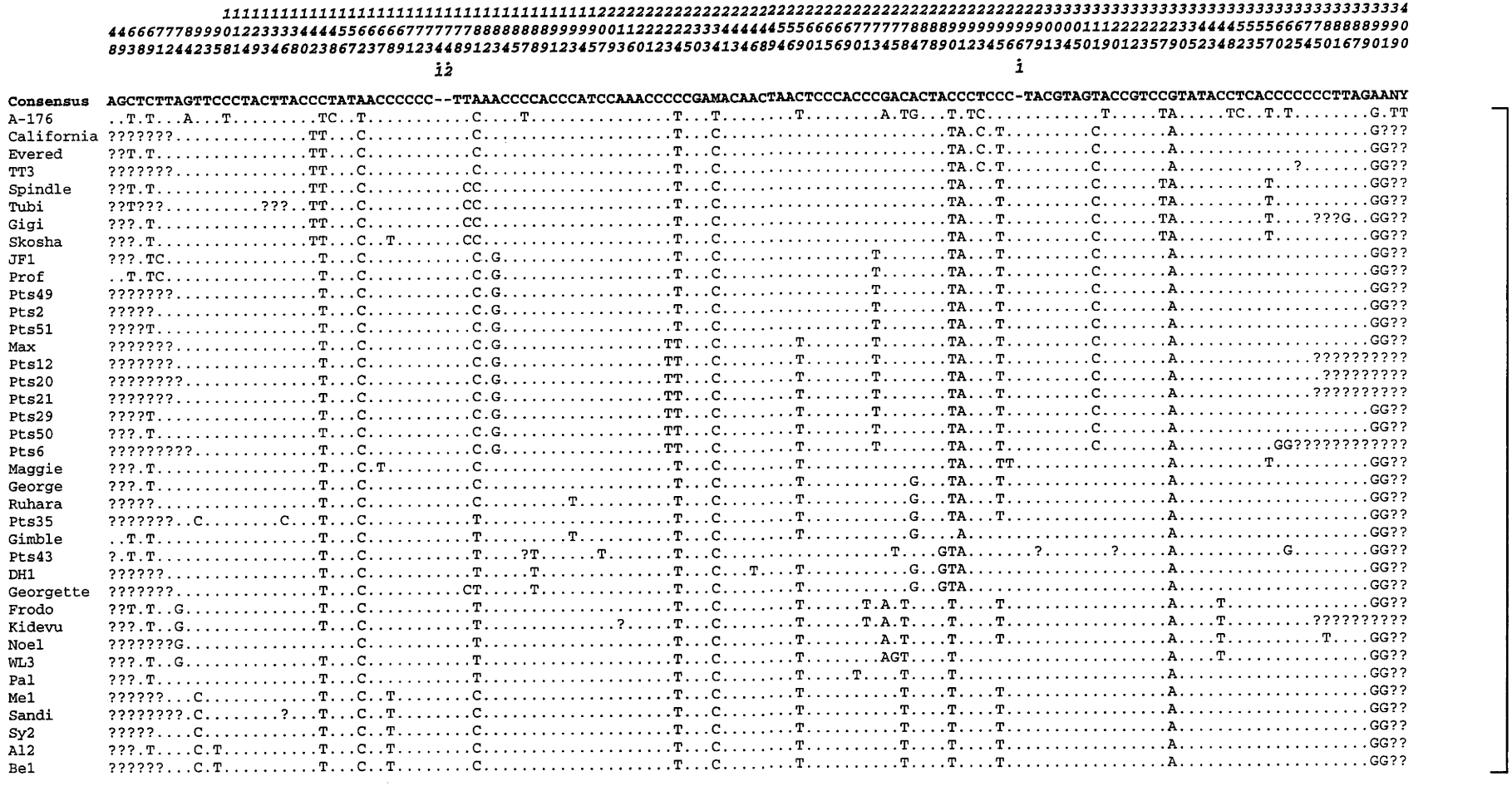
3.4.1 Mitochondrial Control Region Diversity

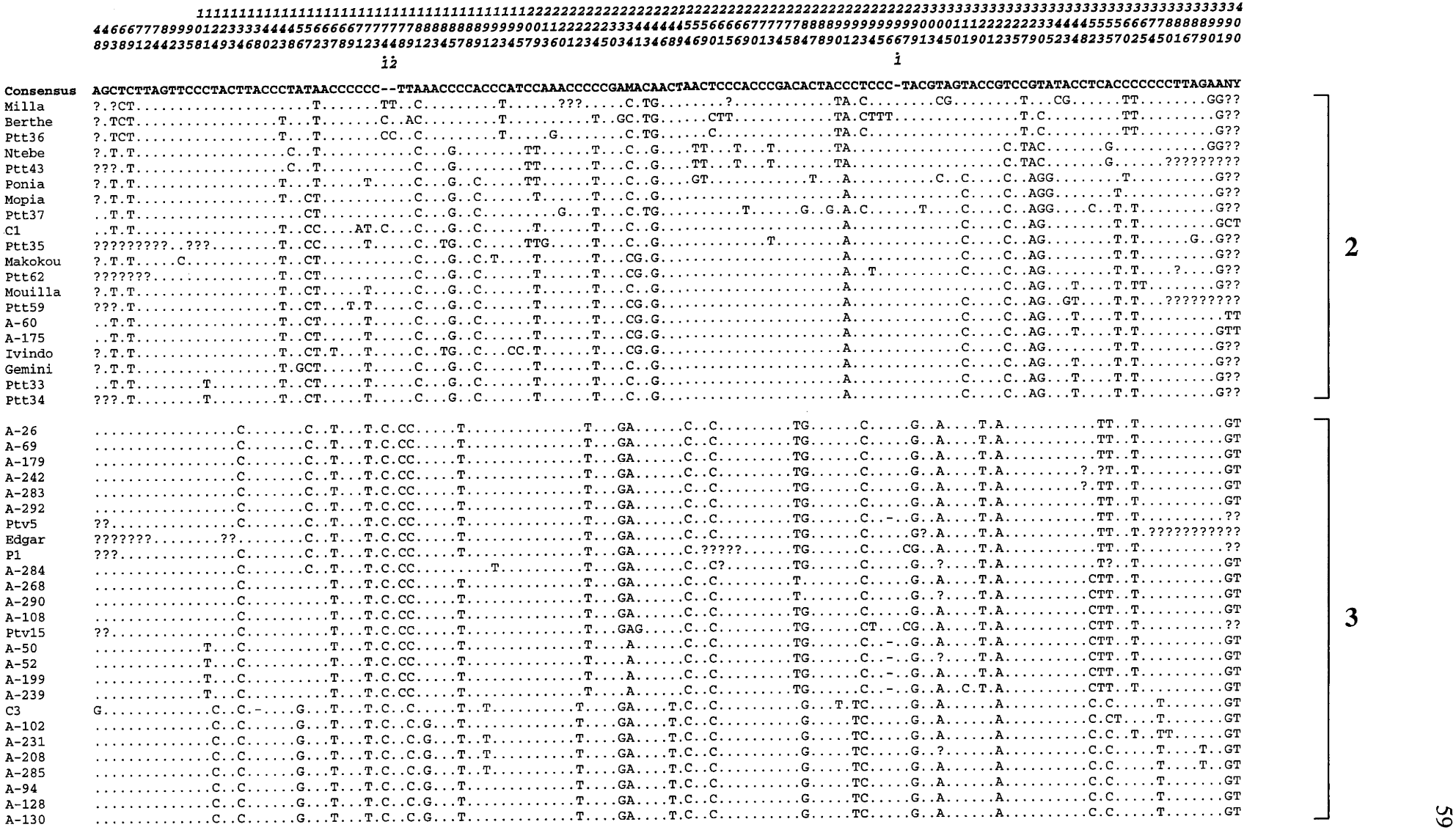
The first hypervariable region of the control region has previously been sequenced for 66 chimpanzees (Kocher and Wilson 1991; Morin *et al.* 1994), the majority of which were of known geographic origin. With the addition of the present 49 sequences, the control region sequences from 115 chimpanzees were aligned (Appendix A). Morin *et al.* (1994) found that chimpanzee control region sequences cluster into three distinct subspecies clades. The bootstrapped NJ tree presented in Figure 3.1 implies that the majority of the sequences in the present study are from the west African subspecies *P. t. verus*, as they cluster with the sequences from this subspecies reported by Morin *et al.* (1994). Two sequences appear to be derived from the central African subspecies *P. t. troglodytes* (A–60 and A–175 in Figure 3.1), and five sequences cannot be ascribed to any of the three subspecies (A–172, A–192, A–197, A–281, and A–176 in Figure 3.1). However, further analysis of the data suggests that A–172, A–192, A–197, and A–281 belong to the newly recognised subclade of western chimpanzees in Nigeria (T. Disotell, personal communication; Gonder *et al.* 1997), and A–176 is from the east African subspecies *P. t. schweinfurthii* (H.-J. Bandelt, personal communication).

FIGURE 3.1 Phylogenetic tree relating the sequences of this study and other published chimpanzee sequences using *P. paniscus* as the outgroup. Sequences from this study have the prefix "A"; sequences obtained from Kocher and Wilson (1991) are C1 to C3; and sequences obtained from Morin *et al.* (1994) are named according to that study. The tree was constructed using the neighbor-joining method (Saitou and Nei 1987) based on Kimura (1980) two-parameter distances. The reliability of each interior branch was tested by 1,000 bootstrap replications, and reliabilities are shown as percentage values next to the major branches. Based on previous characterisation of chimpanzee control region sequences (Morin *et al.* 1994), the three distinct clades (1 to 3) correspond respectively to the three subspecies *P. t. schweinfurthii*, *P. t. troglodytes*, and *P. t. verus*. The majority of the samples in the present study appear to be from the west African subspecies *P. t. verus*, as they fall within clade 3.


A-176
Evered
TT3
California
Spindle
Gigi
Tubi
Skosha
Prof
JF1
Pts49
Pts2
Pts51
Max
Pts29
Pts50
Pts20
Pts21
Pts12
Pts6
Maggie
George
Ruhara
Pts35
Gimble
Pts43
DH1
Georgette
Frodo
Kidevu
Noel
WL3
Pal
Sy2
Be1
Al2
Me1
Sandi
Milla
Berthe
Ptt36
Ptt43
Ntebe
Ponia
Mopia
Ptt37
C1
Ptt35
Makokou
Ptt62
Mouilla
Ptt59
A-60
A-175
Ivindo
Gemini
Ptt33
Ptt34
A-26, A-69, A-179, A-242, A-283, A-292
Ptv5
Edgar
P1
A-284
A-108
Ptv15
A-268, A-290
A-239
A-50, A-52, A-199
C3
A-102
A-231
A-208, A-285
A-94, A-128, A-130
A-33
A-56
A-288
A-137, A-286
NC588
A-90, A-139
A-97
Ptv4
P3
C2
Tobar
A-42, A-89, A-118, A-131, A-136, A-230, A-291
P2a
A-129
A-101
A-282
A-235
A-182
A-62
A-172, A-192, A-197, A-281
P.paniscus
100
44
27
36
94
81
65
92
1
2
3

The sites that vary among these 115 sequences are shown in Figure 3.2. There were 77 distinctive sequence types (missing data excluded) defined by 131 variable nucleotide sites, and five insertions/deletions. Within *P. t. schweinfurthii*, there were 23 sequence types among 38 individuals, and eight sequence types were shared among two, three or six individuals. Within *P. t. troglodytes*, there were 18 sequence types among 20 individuals. Within *P. t. verus*, there were 36 sequence types among 57 individuals, and nine sequence types were shared among two, three, four or six individuals; no sequence types were shared between subspecies. At two sites (positions 16078 and 16166), all chimpanzee sequences lacked the A residue present in the human sequence (Anderson *et al.* 1981) (see Appendix A). Most differences between the sequences (82.5%) result from transition-type mutations. Moreover, transitions between pyrimidines (71.5%) occur much more frequently than those between purines, probably reflecting the low G content in the L-strand of the mitochondrial genome. In this respect the variation is similar to that of human sequences (*e.g.*, Aquadro and Greenberg 1983; Greenberg *et al.* 1983; Vigilant *et al.* 1989; Horai and Hayasaka 1990; Kocher and Wilson 1991; Horai *et al.* 1993; Mountain *et al.* 1995; Watson *et al.* 1996).


Consensus AGCTCTTAGTTCCCTACTTACCCTATAACCCCCC--TTAAACCCCACCCATCCAAACCCCCGAMACAACTAACTCCCACCCGACACTACCCTCCC-TACGTAGTACCGTCCGTATACCTCACCCCCCCTTAGAANY
A-176
California
Evered
TT3
Spindle
Tubi
Gigi
Skosha
JF1
Prof
Pts49
Pts2
Pts51
Max
Pts12
Pts20
Pts21
Pts29
Pts50
Pts6
Maggie
George
Ruhara
Pts35
Gimble
Pts43
DH1
Georgette
Frodo
Kidevu
Noel
WL3
Pal
Mel
Sandi
Sy2
Al2
Be1
1


Consensus AGCTCTTAGTTCCCTACTTACCCTATAACCCCCC--TTAAACCCCACCCATCCAAACCCCCGAMACAACTAACTCCCACCCGACACTACCCTCCC-TACGTAGTACCGTCCGTATACCTCACCCCCCCTTAGAANY
Milla
Berthe
Ptt36
Ntebe
Ptt43
Ponia
Mopia
Ptt37
C1
Ptt35
Makokou
Ptt62
Mouilla
Ptt59
A-60
A-175
Ivindo
Gemini
Ptt33
Ptt34
A-26
A-69
A-179
A-242
A-283
A-292
Ptv5
Edgar
P1
A-284
A-268
A-290
A-108
Ptv15
A-50
A-52
A-199
A-239
C3
A-102
A-231
A-208
A-285
A-94
A-128
A-130
2
3

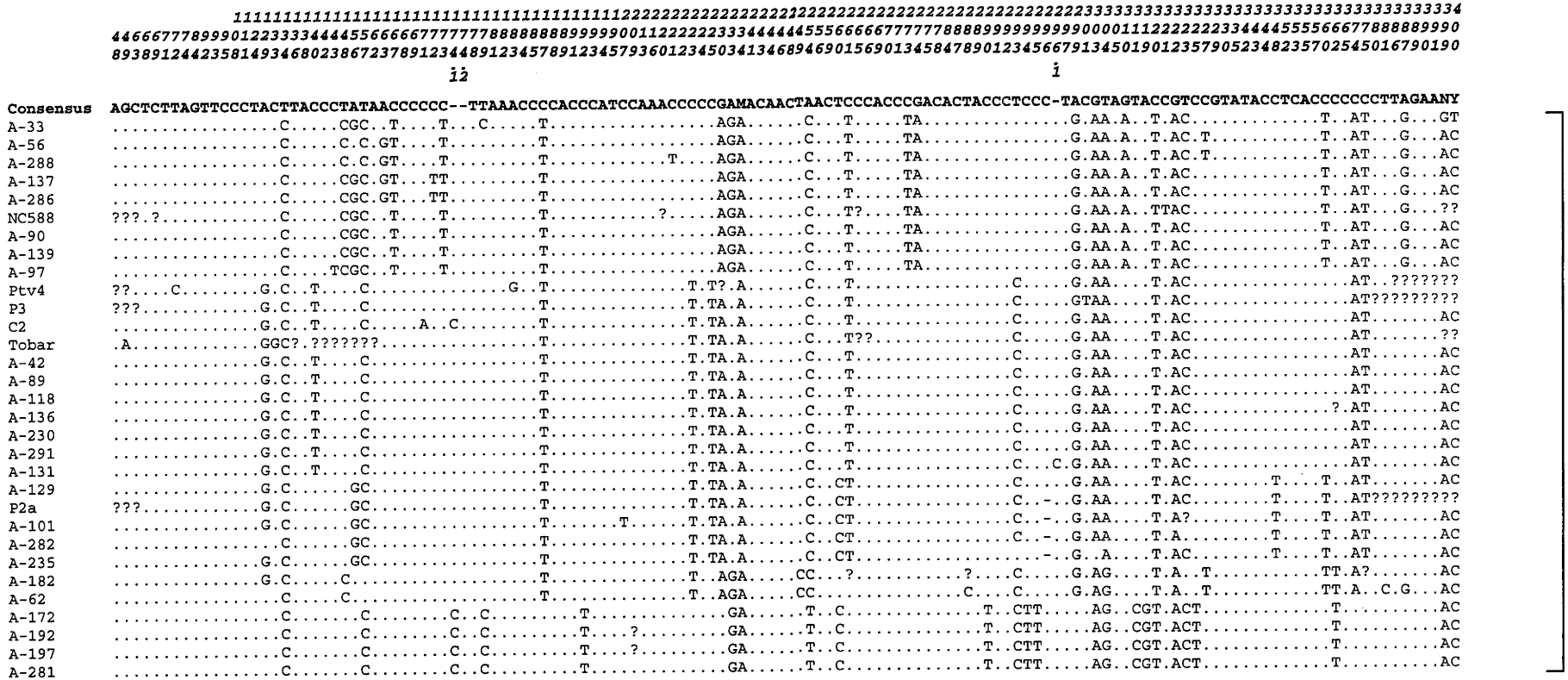


FIGURE 3.2 Variable nucleotide sites in a 377 bp segment of the mitochondrial control region for 115 chimpanzees. A total of 131 variable sites and five insertions/deletions were observed; shown as differences from a chimpanzee consensus sequence. Nucleotide position numbers correspond to the last three digits in the numbering system of Anderson *et al.* (1981) (*e.g.*, 48 = 16,048) and numbers followed by a decimal point indicate additional nucleotides not found in this sequence. Dots (.) indicate sequence identity, dashes (-) denote insertions/deletions, ? denotes missing data and the IUB single-letter codes for nucleotide bases have been used for ambiguous bases. Sequences reported in this study have the prefix "A"; sequences obtained from Kocher and Wilson (1991) are C1 to C3; and sequences obtained from Morin *et al.* (1994) are named according to that study. Individuals have been arranged by sequence type and subspecies separations match those in Figure 3.1.

The mean number of nucleotide differences per site (π) between sequences was estimated as previously outlined in Section 3.3.3. Table 3.1 shows uncorrected and corrected mean and maximum pairwise distances for all 115 chimpanzee sequences, and for each of the three subspecies. Mismatch distributions, frequency distributions of pairwise distances between individuals (Rogers and Harpending 1992), for the three chimpanzee subspecies are shown in Figure 3.3. The distributions are significantly different from the Poisson distributions predicted for the observed mean values using a chi-square test ($P < 0.001$). The distributions for *P. t. verus* and *P. t. troglodytes* are clearly multimodal (Figure 3.3 *A* and *B*), a pattern consistent with stable populations (Slatkin and Hudson 1991; Rogers and Harpending 1992). The *P. t. schweinfurthii* distribution is approximately unimodal (Figure 3.3 *C*), which Rogers and Jorde (1995) have interpreted as indicating that this subspecies has experienced a recent population expansion. The uncorrected mean (0.024 ± 0.013) and maximum (0.072) pairwise distances in *P. t. schweinfurthii* are also lower than in *P. t. verus* (mean: 0.051 ± 0.025; maximum: 0.093) and *P. t. troglodytes* (mean: 0.040 ± 0.021; maximum: 0.091). Both mismatch distribution analysis and the overall pattern of low genetic variability in eastern chimpanzees supports the hypothesis that the subspecies has responded to recent climatically induced changes (20,000–60,000 years ago) in the distribution of eastern African forests (Goldberg and Ruvolo 1997).

The mean and maximum pairwise distances in a diverse sample of 1,554 humans are 0.020 ± 0.011 and 0.067, respectively (Watson 1996). In a sample of 389 Africans, who contain the most divergent of the human mtDNA lineages (Cann *et al.* 1987*a*; Vigilant *et al.* 1989, 1991; Horai *et al.* 1993), they are 0.025 ± 0.013 and 0.062, respectively (Watson 1996). Thus, the nucleotide distances within chimpanzee subspecies are the same as or greater than the distances within the entire human species, and the overall distances in chimpanzees (mean: 0.075 ± 0.037; maximum: 0.139) are substantially greater than in humans. This is consistent with the greater levels of variation observed in studies of restriction site polymorphisms (Ferris *et al.* 1981) and DNA sequences (see Chapter 4; Morin *et al.* 1994; Ruvolo *et al.* 1994; Nachman *et al.* 1996; Wise *et al.* 1998) in the mitochondrial genomes of various ape species.

TABLE 3.1 Overview of the mitochondrial control region variation in chimpanzees.

Species/ subspecies	*n*	lineages	*S*	*p*-distance		Kimura 2-parameter		Tamura-Nei	
				$\pi \pm$ s.d.	max. π	$\pi \pm$ s.d.	max. π	$\pi \pm$ s.d.	max. π
Pan troglodytes	115	77	136	0.075 ± 0.037	0.139	0.083 ± 0.040	0.163	0.103 ± 0.050	0.227
P. t. schweinfurthii	38	23	50	0.024 ±0.013	0.072	0.025 ± 0.013	0.076	0.028 ± 0.014	0.094
P. t. troglodytes	20	18	69	0.040 ± 0.021	0.091	0.042 ± 0.022	0.099	0.050 ± 0.037	0.135
P. t. verus	57	36	80	0.051 ± 0.025	0.093	0.054 ± 0.027	0.102	0.065 ± 0.032	0.134

NOTE.—*n* is the number of individuals examined; lineages are the number of different sequence types (no sequence types were shared between subspecies); *S* is the number of segregating (polymorphic) sites from nucleotide positions 16024 to 16400 (including insertion/deletion events); π is the mean pairwise distance; s.d. is the standard deviation; and max. π is the maximum pairwise distance. Mean pairwise distances and their standard deviations were calculated as outlined in Section 3.3.3. The standard deviation estimates include both sampling and stochastic variance.

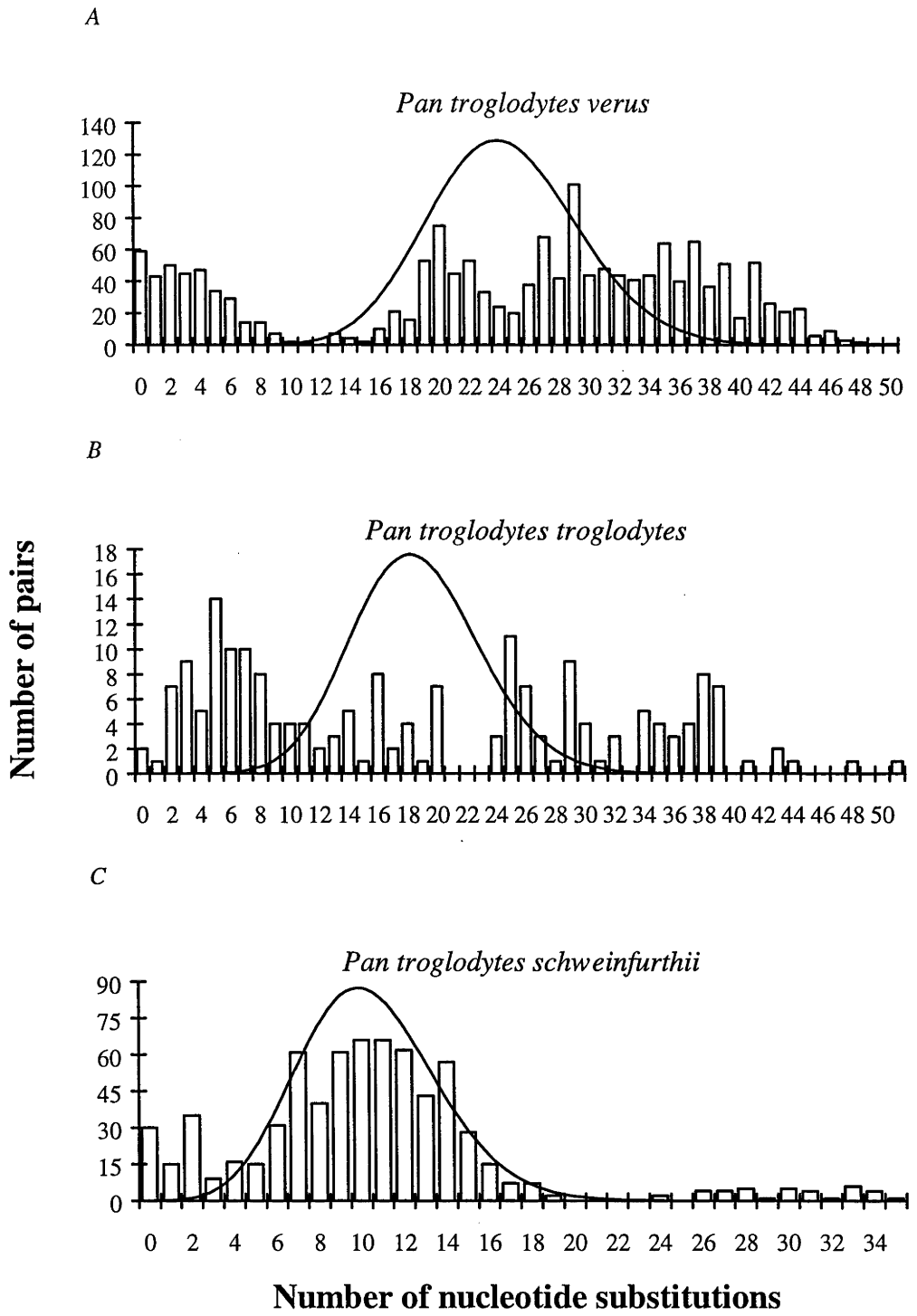


FIGURE 3.3 Pairwise distributions of nucleotide distances (Tamura and Nei 1993) in three chimpanzee subspecies. *A*, Fifty-seven west African chimpanzees. *B*, Twenty central African chimpanzees. *C*, Thirty-eight east African chimpanzees. The lines show the expected Poisson distributions having the same means as the observed distributions.

3.4.2 Nuclear Genome Diversity

A comparison of levels of heterozygosity in humans and chimpanzees at eight tandem repeat loci analysed in the present study (refer Section 2.5 and Section 3.3.4) and 60 loci for which variation in the two species had been described elsewhere is presented in Table 3.2. The heterozygosity of human loci is significantly higher than that of chimpanzees by the nonparametric Wilcoxon's signed-ranks test ($z = -2.99$, $P < 0.005$) (Seigel 1956).

All but two of the loci analysed were already known to be polymorphic in humans. The possibility of ascertainment bias resulting in the relative overestimation of heterozygosity in humans must therefore be considered. Although bias in heterozygosity estimates cannot be completely ruled out there are two reasons why it is unlikely. First, the mean and the range of heterozygosities for the analysed loci are similar in humans to those found in a sample of randomly selected loci (Hudson *et al.* 1992). Thus, they do not appear to be biased in favour of loci with high heterozygosity. Second, ascertainment bias will only result in a pronounced estimate bias when heterozygosity in the source species is less than approximately 0.4; the bias is negligible when heterozygosity exceeds approximately 0.5 (Rogers and Jorde 1996). Thus, the tandem repeat loci analysed in the present study that are polymorphic in both humans and chimpanzees should be essentially free of the effect of ascertainment bias because of their high heterozygosities in humans. Bias may still occur, however, if loci are selected that are highly polymorphic in humans but essentially monomorphic in chimpanzees because of base substitutions that interfere with the generation of repeat-length polymorphism (Crouau-Roy *et al.* 1996). To avoid such bias, the eight loci that were highly polymorphic in humans but monomorphic, or nearly monomorphic, in chimpanzees were excluded from analysis.

The higher level of diversity at tandem repeat loci in humans is consistent with data from other kinds of loci. Chimpanzees also have fewer polymorphic red cell enzyme and serum protein loci, and only 0.1–0.2 of the average heterozygosity (H) found at these loci in humans (King and Wilson 1975; Bruce and Ayala 1979). There is a possibility of ascertainment bias in these comparisons since loci may have been selected for analysis because they are polymorphic in humans. However, Harris and Hopkinson (1972) obtained an estimate of $H = 0.072$ in Europeans for 26 enzyme-coding loci chosen

without any prior knowledge of polymorphism. This was very similar to the estimate of 0.065 obtained for 45 other enzyme-coding loci from data compiled from the literature, indicating that, at least for enzyme-coding loci, the heterozygosity estimates in humans are not biased. In later studies (Nei and Roychoudhury 1974*a*, 1982), *H* was estimated as 0.10–0.14, the difference being attributed to the inclusion of non-enzyme-coding loci in the latter studies, which are generally more polymorphic than enzyme-coding loci. The *H* values obtained for chimpanzees include data on non-enzyme-coding loci, thus the human values that are appropriate for comparison with chimpanzees are 0.10–0.14, reported by Nei and Roychoudhury (1974*a*, 1982). However, the more conservative estimate of 0.067 (Harris and Hopkinson 1972) is also substantially greater than the estimates (0.01–0.02) for chimpanzees (King and Wilson 1975; Bruce and Ayala 1979). In addition, of five minisatellite loci compared between humans and chimpanzees (Ely *et al.* 1992), three were polymorphic in both species, with average heterozygosities of 0.72 and 0.61 in humans and chimpanzees, respectively. There are also indications of lower levels of variation in chimpanzees at other minisatellite loci (Wolff *et al.* 1991) and at Major Histocompatibility Complex loci (Kenter *et al.* 1992; Ayala and Escalante 1996).

TABLE 3.2 Genetic diversity at 68 tandem repeat loci in humans and chimpanzees.

Locus	Human			Chimpanzee		
	n	No. of Alleles	Hetero-zygosity	*n*	No. of Alleles	Hetero-zygosity
D13S71[a]	138	5	0.747	186	3	0.187
D13S118[a]	154	8	0.727	168	7	0.738
D13S121[a]	156	8	0.772	156	14	0.883
D13S122[a]	154	12	0.838	178	5	0.628
D13S124[a]	156	6	0.673	178	8	0.824
D13S193[a]	152	10	0.745	182	13	0.723
D13S197[a]	154	22	0.880[b]	108	2	0.018[b]
FES[c]	48	6	0.766	90	4	0.506
MBP1[c]	72	6	0.801	90	7	0.526
MBP2[c]	72	7	0.811	90	7	0.657
SE33[c]	78	21	0.942	90	3	0.364
TH01[c]	1280	8	0.790	90	3	0.293
VWF[c]	200	7	0.734	90	8	0.667
Mfd 3[d]	218	10	0.748	29	10	0.885
Mfd 32[d]	220	11	0.708	32	6	0.825
Mfd 38[d]	220	14	0.844	32	16	0.928
Mfd 59[d]	225	13	0.879	32	14	0.896
Mfd 75[d]	219	15	0.883	32	7	0.719
Mfd 104[d]	182	15	0.837	32	15	0.928
Mfd 139[d]	213	20	0.797	32	6	0.677
Mfd 142[d]	204	10	0.758	32	4	0.708
D4S174[e]	100	11	0.826	56	17	0.974
D4S190[e]	100	8	0.804[b]	55	6	0.266[b]
D4S230[e]	100	12	0.824	56	12	0.886
D4S391[e]	100	13	0.927	56	13	0.975
D4S404[e]	100	5	0.701[b]	56	3	0.107[b]
D4S405[e]	100	9	0.866	56	8	0.797

Locus	Human			Chimpanzee		
	n	No. of Alleles	Hetero-zygosity	*n*	No. of Alleles	Hetero-zygosity
D4S418[e]	100	9	0.824	56	7	0.886
D4S419[e]	100	6	0.727	54	9	0.798
D4S425[e]	100	7	0.715	56	7	0.926
D4S551[e]	100	12	0.712	56	11	0.891
D4S616[e]	100	10	0.828	56	12	0.752
D4S885[e]	100	8	0.748[b]	56	2	0.132[b]
D2S119[f]	14	8	1.000	16	7	1.000
D2S123[f]	16	7	0.933	16	3	0.133
D2S391[f]	16	4	0.533	16	1	0.000
D3S1029[f]	16	6	0.933	16	5	0.267
D3S1038[f]	16	6	0.800	16	3	0.267
D3S1076[f]	16	5	0.667	16	6	0.667
D3S1298[f]	16	9	0.800	16	5	0.133
D3S1561[f]	16	8	0.933	16	4	0.267
D9S66[f]	16	9	1.000	8	3	0.286
D9S104[f]	16	6	0.400	16	3	0.800
D9S122[f]	16	6	0.933[b]	16	1	0.000[b]
D9S150[f]	16	6	0.800[b]	16	1	0.000[b]
D9S298[f]	16	4	0.533	16	3	0.933
DXS3[f]	8	4	1.000	16	4	0.533
DXS207[f]	8	4	1.000	16	7	0.800
DXS228[f]	8	6	1.000[b]	16	2	0.000[b]
DXS453[f]	8	5	0.857	16	6	0.667
DXS1110[f]	8	5	0.571	16	7	0.800
DYS II[f]	8	5	0.571	16	3	0.800
CYBB[f]	8	7	0.857	16	4	1.000
DBH[f]	14	3	0.769	16	3	0.400
H1[f]	8	3	0.571	16	6	0.667

Locus	Human			Chimpanzee		
	n	No. of Alleles	Hetero-zygosity	*n*	No. of Alleles	Hetero-zygosity
L128[f]	6	5	1.000	16	5	0.533
MAOB[f]	8	7	1.000	16	3	0.400
PFC[f]	8	9	1.000	14	5	1.000
STR44[f]	8	8	1.000	16	5	0.800
X75b[f]	16	6	0.933	16	3	0.400
COL2A1[g]	32	5	0.836	100	5	0.505
D7S460[g]	48	?	0.970[b]	78	1	0.000[b]
D8S342[g]	66	10	0.619	96	5	0.566
MYCN[g]	118	3	0.424	102	3	0.613
RENA4[g]	1250	6	0.399	104	2	0.308
TH04[g]	70	5	0.791	102	2	0.357
BG01[h]	98	7	0.717	104	5	0.444
EG01[h]	102	2	0.040	102	1	0.000
Mean		8.033	0.776		6.383	0.630
Standard error		0.498	0.023		0.501	0.035

NOTE.—*n* is the number of chromosomes examined. If two related individuals shared one allele at a locus, one copy was excluded, causing allele frequencies (and thereby the expected heterozygosity) to be based on an odd number of chromosomes (J. C. Garza, personal communication). The heterozygosity of each locus was estimated as outlined in Section 3.3.4. The mean heterozygosity and its standard error were calculated according to Nei and Roychoudhury (1974*b*).

Loci: [a] Deka *et al.* (1994); [b] Data excluded from analysis due to negligible variation in chimpanzees despite high heterozygosity in humans; [c] Pascall *et al.* (1994) and references therein; [d] Garza *et al.* (1995); [e] Crouau-Roy *et al.* (1996); [f] D. C. Rubinsztein, personal communication; [g] Only chimpanzees analysed in the present study. Data sources for humans are as in Table 2.3; [h] Both humans and chimpanzees analysed in the present study.

3.5 DISCUSSION

The consistency of tandem repeat loci and other kinds of loci in showing a higher level of nuclear genome diversity in humans than in chimpanzees implies that the nuclear genome as a whole is more variable in humans. In contrast, the mitochondrial genome is less variable in humans than in chimpanzees. However, the discrepancy between the levels of mitochondrial and nuclear genome diversity in humans and chimpanzees found in this study is based on samples that were collected rather differently in the two species. The possibility that the result is a sampling artefact needs to be considered.

The microsatellite diversity was estimated largely from samples of European humans and west African chimpanzees, although in both species individuals from other geographical areas were analysed for some loci. European humans do not appear to have higher levels of microsatellite diversity than humans from other regions (Jorde *et al.* 1995; Richards *et al.* 1996). It is possible that west African chimpanzees have lower levels of nuclear genome diversity than chimpanzees from other regions (although they do not have lower levels of mitochondrial genome diversity; refer Table 3.1).

As previously discussed in Section 3.4.1, mitochondrial diversity is lower in humans regardless of whether comparisons are made between samples collected from one geographical region in both species (*e.g.*, west Africa in chimpanzees and Africa in humans), or across the geographical range of both species. Thus, this difference does not appear to be a sampling artefact.

The discrepancy could be a consequence of one or more factors that differ between humans and chimpanzees. To investigate this further it would be useful to know if either one of the species has an unusual π/H ratio. Table 3.3 compares the levels of nuclear and mitochondrial genome diversity in humans and chimpanzees with those in six other catarrhine taxa. Nuclear genome diversity was estimated as H for protein-coding loci. Mitochondrial genome diversity was estimated as either π_1, the mean pairwise distance between control region sequences, or, in the cases where sequence data were not available, π_2, the mean pairwise distances of the entire mitochondrial genome based on

restriction maps. The total variance which incorporates the sampling variance as well as other stochastic factors (Nei 1987) is very large so that the significance of any differences are not readily apparent. Also, the estimates π_1, π_2, and H may not be directly comparable because: (i) The exact region of the mitochondrial genome sequenced varies among studies. (ii) The number and type of restriction endonucleases used varies among studies. (iii) Different protein-coding loci are used in different studies. With these caveats in mind, it can be seen that the human π_1/H and π_2/H ratios are both substantially less than those of all other species. In contrast, the chimpanzee ratios are similar to those of the other non-human species (Table 3.3). The discrepancy thus appears to be a result of an unusually low π/H ratio in humans rather than an unusually high ratio in chimpanzees.

The estimate presented here of the difference in π/H ratios between humans and chimpanzees is conservative. European humans have unusually low levels of mitochondrial genome diversity ($\pi = 0.013 \pm 0.007$; Watson 1996), so by restricting the nuclear genome comparison to a largely European human sample but extending the mitochondrial genome comparison to include individuals from other regions, the extent of the difference in π/H ratios between the species may have been underestimated. If the comparison were restricted largely to European humans ($\pi_1/H = 0.19$) and west African chimpanzees ($\pi_1/H = 2.43$), the discrepancy in the ratios between these representative groups of the two species would be 12.8-fold, compared to 11.9-fold (Table 3.3).

Some researchers believe that a small, long-term average N_e explains both mitochondrial and nuclear genome diversity in humans (Takahata 1993*a*; Hammer 1995), implying that the π/H ratio is not unusual. However, predictions about long-term average N_e only hold if human population size has been constant since our divergence from chimpanzees some 5 Ma ago. Arcaeological evidence suggests the last one million years or so is marked by continued range expansions and changes in population size (Gamble 1994)

TABLE 3.3 Mitochondrial and nuclear genome diversity in catarrhines.

Species	π_1	π_2	H	π_1/H	π_2/H
Homo sapiens	0.020[a] ± 0.011	0.0032[d]	0.067[h] ± 0.02	0.30	0.05
Pan troglodytes	0.075[b] ± 0.037	0.0133[e]	0.021[i] ± 0.01	3.57	0.63
Gorilla gorilla	0.099[c] ± 0.050	0.0055[e]	0.049[j] ± 0.03	2.02	0.11
Pan paniscus	–	0.0100[e]	0.022[j] ± 0.02	–	0.45
Pongo pygmaeus abelii	–	0.0210[e]	0.048[j] ± 0.03	–	0.44
Pongo pygmaeus pygmaeus	–	0.0050[e]	0.025[j] ± 0.03	–	0.20
Macaca fuscata	–	0.0132[f]	0.013[k] ± 0.001	–	1.02
Macaca fascicularis	–	0.0410[g]	0.096[l]	–	0.43

NOTE.—π_1, mean pairwise distances based on nucleotide sequences from a 300–400 bp segment of the first hypervariable region of the mitochondrial control region. Distances between each pair of sequences were computed using an uncorrected proportional distance, and the mean and standard deviation were calculated as outlined in Section 3.3.3. The standard deviation incorporates the variance due to stochastic factors and sampling errors. π_2, mean pairwise distances based on restriction maps of the entire mitochondrial genome. H, average heterozygosity for protein-coding loci weighted according to sample size using Equation 2.5 in Section 2.6.3.3. The standard error of H was calculated according to Nei and Roychoudhury (1974*b*). H is estimated in all species except humans using data on both enzyme-coding and non-enzyme-coding loci. The human estimate is based only on the less variable enzyme-coding loci since this estimate appears not to be biased by inclusion of a disproportionate number of polymorphic loci (Harris and Hopkinson 1972). The estimate is conservative, and inclusion of data for non-enzyme-coding loci gives human H estimates in the range 0.10 to 0.14 (Nei and Roychoudhury 1974*a*, 1982).

References are as follows: [a] 1,554 worldwide humans (Watson 1996); [b] Data combined from the present study and Morin *et al.* (1994); [c] Garner and Ryder (1996); [d] Cann *et al.* (1987*a*); [e] Ferris *et al.* (1981); [f] Hayasaka *et al.* (1986); [g] Harihara *et al.* (1988); [h] Harris and Hopkinson (1972); [i] King and Wilson (1975); [j] Bruce and Ayala (1979); [k] Nozawa *et al.* (1982); and [l] Nei and Graur (1984).

A low human π/H ratio could result from an unusual mutation rate, either an unusually high nuclear genome mutation rate or an unusually low mitochondrial genome mutation rate. Mutation rate differences, however, would result in substitution rate differences in the human lineage compared with the chimpanzee lineage, and no such differences are apparent (Sibley and Ahlquist 1987; Horai *et al.* 1992; Easteal *et al.* 1995). It has been suggested that microsatellite mutation rates are higher in humans than in chimpanzees (Rubinsztein *et al.* 1995*b*), possibly related to differences in allele length (Rubinsztein *et al.* 1995*a*). However, differences in both allele length and mutation rate have been questioned by Ellegren *et al.* (1995), who suggest that the apparent allele length difference may be due to ascertainment bias. This issue has yet to be resolved. Evidence for a mutation rate difference is based on discrepancies of the estimated and expected genetic distances between humans and chimpanzees and between two human groups (Europeans and sub-Saharan Africans). No estimate was made of divergence within chimpanzees. The validity of the suggested mutation rate difference depends on many factors including the extent of divergence within chimpanzees, the relationship between genetic distance and separation time over a period of several million years, the stochastic error in the estimate of genetic distance, the assumed divergence time of humans and chimpanzees, and the assumed divergence time of sub-Saharan Africans and Europeans. Although the possibility of differences in mutation rates cannot be excluded, further investigation is needed before it is clearly demonstrated.

A number of other factors could explain the low human π/H ratio. For example, π/H will decrease as the ratio of males to females increases—π depends only on the effective number of females, while H depends on the effective number of both males and females (Birky *et al.* 1983). However, the relationship is asymptotic and even extreme sex ratio disparity is unlikely to explain the degree of anomaly in the observed π/H ratio. Similarly, in a subdivided population, if female migration rates greatly exceed those of males some reduction in π/H will occur (Birky *et al.* 1989). However, male philopatry and female dispersal are known to occur in chimpanzees (Morin *et al.* 1994). Thus, in order to explain the difference in π/H ratio between humans and chimpanzees, a human bias in female migration would have to be extreme.

The anomalous π/H ratio in humans may reflect either diversifying selection acting to increase variation in the nuclear genome and/or directional selection acting to reduce variation in the mitochondrial genome. The nuclear genome diversity estimates are based on many independent loci whereas the mitochondrial genome is a single linkage group. Natural selection on a mitochondrial coding gene will therefore exert a substantial genetic "hitchhiking" effect, even on polymorphisms in the noncoding control region. On the other hand, increased nuclear genome diversity would only result from selection acting on many different loci independently. It is unlikely that balancing selection has acted to enhance nuclear diversity at tandem repeat loci as well as protein coding loci. Thus, the mitochondrial genome is the most likely candidate for natural selection, and this is further investigated in Chapter 4.

The results presented here establish the generality of the discrepancy, and contribute further to its quantification. This provides the basis for a detailed investigation of the possible ways in which these factors might have interacted to produce the discrepancy. This will have important implications with respect to the current dynamics of sequence evolution in human mtDNA.

Chapter 4

PATTERNS OF MITOCHONDRIAL VARIATION WITHIN AND BETWEEN HUMANS AND CHIMPANZEES

4.1 ABSTRACT

To test whether patterns of mtDNA variation are consistent with a neutral model of molecular evolution, nucleotide sequences were determined for the 1041 base pairs of the NADH dehydrogenase subunit 2 (*ND2*) gene and the 1140 base pairs of the cytochrome *b* (*cyt b*) gene in 20 geographically diverse humans and 20 common chimpanzees. Contingency tests of neutrality were performed using four mutational categories for the ND2 molecule: synonymous and nonsynonymous mutations in the transmembrane regions, and synonymous and nonsynonymous mutations in the surface regions. The surface regions of the cytochrome *b* molecule were further divided into internal and external surfaces resulting in six mutational categories. Three topological mutational categories were also used: intraspecific tips, intraspecific interiors, and interspecific fixed differences. Both the *ND2* data and the *cyt b* data reveal a higher ratio of nonsynonymous to synonymous nucleotide differences within humans than is seen between species, although for *cyt b* the difference is not significant. The nested contingency analyses of the *ND2* data reveal a significantly greater number of nonsynonymous polymorphisms (or deficiency of synonymous polymorphisms) within human transmembrane regions than expected based on interspecific comparisons, and they are inconsistent with a neutral equilibrium model. This pattern of nonsynonymous and synonymous variation is not seen within chimpanzees. Other statistical tests of neutrality, such as Tajima's *D* test and the *D* and *F* tests proposed by Fu and Li, indicate an excess of low frequency polymorphisms in the human data, but not in the chimpanzee data. The combined test results are consistent with recent directional selection or background selection against slightly deleterious mutations in the human mitochondrial genome. The analyses further support the idea that mitochondrial genome evolution is governed by selective forces that have the potential to affect its use as a "neutral" marker in evolutionary and population genetic studies.

4.2 INTRODUCTION

The neutral theory makes a number of straightforward predictions, and thus serves as a useful null hypothesis for studies of genetic variation within and between species. One such prediction is that the amount of nucleotide polymorphism within a species will be correlated with the amount of divergence between species, and statistical tests which are essentially goodness-of-fit tests to neutrality have been devised on this principle (*e.g.*, Hudson *et al.* 1987). An additional prediction of the neutral theory, formulated into a test by McDonald and Kreitman (1991), is that the ratio of amino acid replacement (nonsynonymous) to silent (synonymous) nucleotide differences will be the same within and between species.

Mitochondrial DNA is widely used as a marker in evolutionary and population genetic studies, and is generally assumed to evolve according to a neutral model of molecular evolution. This assumption may be important for such things as measuring gene flow (Slatkin 1985), estimating effective population size (Wilson *et al.* 1985), detecting population subdivision (Avise *et al.* 1987), and dating times of species divergence or historical events within a species using a molecular clock (Brown 1980; Cann *et al.* 1987*a*; Vigilant *et al.* 1991). The extent of hitchhiking of neutral mutations in response to selection on another part of the genome depends on the recombination distance from the site under selection (Maynard-Smith and Haigh 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992). In the extreme case of no genetic recombination, such as in mtDNA, the whole genome is a single completely linked entity. Any selective force acting at one site will equally affect the history of the whole molecule. Thus, the selective fixation of an advantageous mutation, for example, will lead to the concomitant fixation of all other polymorphisms in the genome.

A number of studies have compared patterns of RFLP variation in human mtDNA to neutral predictions (*e.g.*, Johnson *et al.* 1983; Whittam *et al.* 1986; Excoffier 1990; Merriwether *et al.* 1991) and found fewer intermediate frequency polymorphisms than expected using Watterson's (1978) test of homozygosity and/or Tajima's (1989*a*) test. While these findings are inconsistent with neutral expectations, it is unclear whether the

deviations arise from changing population sizes, recent natural selection, or both. Rogers and Harpending (1992) studied the distribution of pairwise nucleotide differences for human mitochondrial data and found that the distribution does not conform to a neutral equilibrium model. They suggest that the results fit well with a rapid expansion in population size or a population bottleneck. A bottleneck could well have been the result of a selective sweep of a mtDNA type rather than an actual population size reduction. In these studies, departures from the neutral model can be explained by a variety of processes including selection. Evidence for selection in mtDNA comes from more recent studies utilising the McDonald and Kreitman (1991) approach (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Templeton 1996). Common to all of these studies was the finding of higher ratios of nonsynonymous to synonymous nucleotide differences within species than between species either for all or part of the genes in question.

In humans and chimpanzees previous studies have involved DNA sequence data from the 345 bp *ND3* gene (Nachman *et al.* 1996) and the 783 bp *COII* gene (Templeton 1996) using limited sample sizes (particularly chimpanzees). To investigate neutral predictions further, nucleotide sequences were determined for the NADH dehydrogenase subunit 2 (*ND2*) gene and the cytochrome *b* (*cyt b*) gene in 20 geographically diverse humans and 20 common chimpanzees. The patterns of variation within species were compared to the patterns between species, utilising a simple contingency test of neutrality (Templeton 1987, 1996; McDonald and Kreitman 1991). Departures from neutrality were also investigated using Tajima's (1989*a*) *D* test (hereafter referred to as D_T), and Fu and Li's (1993) *D* (hereafter referred to as D_F) and *F* tests.

4.3 METHODS

4.3.1 Samples

Twenty human samples from indigenous populations of Africa, Europe, Asia and Australia were analysed for each of the *ND2* and *cyt b* genes. Fourteen of these samples were analysed for both genes, and six samples were unique to each gene. Sample details are shown in Table 4.1. Twenty common chimpanzee samples of unknown geographic origin were also analysed for both genes. Previous analysis of mitochondrial control region sequences (see Section 3.4.1; Wise *et al.* 1997), however, indicates that all but two of the individuals (A-175 and A-176) included in the present study are from the west African subspecies *P. t. verus*. Furthermore, individual A-281 appears to belong to the newly recognised subclade of western chimpanzees in Nigeria (T. Disotell, personal communication; Gonder *et al.* 1997).

4.3.2 Sequence Data

DNA sequences encompassing the 1041 bp *ND2* gene (positions 4470 to 5510 in the numbering system of Anderson *et al.* 1981) and the 1140 bp *cyt b* gene (positions 14747 to 15886) were obtained as outlined in Chapter 2. Consensus sequences were produced by aligning forward- and reverse-complement sequences from the same individual in the SeqEd™ DNA Sequence Editor (ABI) program. DDBJ/EMBL/Genbank accession numbers for the *ND2* sequences reported here are AF014882–AF014921; and those for the *cyt b* sequences are AF042500–AF042539. Previously published human (Anderson *et al.* 1981), chimpanzee, bonobo and gorilla (Horai *et al.* 1995) sequences were used in the comparative analyses. Sequence alignment was performed manually using Genetic Data Environment (GDE) 2.2 (Smith *et al.* 1994). Appendix B and Appendix C show the alignments of the *ND2* and *cyt b* genes, respectively, for humans, chimpanzees and gorilla.

TABLE 4.1 Details of the human samples analysed for the *ND2* and *cyt b* genes.

Numbering in this study	Geographic origin	Laboratory designation	Region analysed
AFR1	Bantu	1.41	*ND2, cyt b*
AFR2	Bantu	13.40	*ND2, cyt b*
AFR3	Bantu	17.40	*ND2*
AFR4	Bantu	25.40	*ND2*
AFR5	Bantu	701.40	*ND2*
AFR6	Bantu	Afr1	*cyt b*
AFR7	Bantu	Afr2	*cyt b*
AFR8	Bantu	Afr3	*cyt b*
EUR1	Anglo-Celt	CG	*ND2, cyt b*
EUR2	Anglo-Celt	LC	*ND2, cyt b*
EUR3	Anglo-Celt	MN	*ND2, cyt b*
EUR4	Anglo-Celt	SS	*ND2, cyt b*
EUR5	Anglo-Celt	VQ	*ND2, cyt b*
ASN1	Cantonese	HK3845	*ND2, cyt b*
ASN2	Cantonese	HK3853	*ND2, cyt b*
ASN3	Cantonese	HK3863	*ND2, cyt b*
ASN4	Cantonese	HK3918	*ND2*
ASN5	Cantonese	HK3968	*ND2*
ASN6	Cantonese	HK4110	*cyt b*
ASN7	Cantonese	HK4114	*cyt b*
AUS1	Aboriginal Australian	MOW4	*ND2, cyt b*
AUS2	Aboriginal Australian	MOW6	*ND2*
AUS3	Aboriginal Australian	MOW8	*ND2, cytb*
AUS4	Aboriginal Australian	MOW11	*ND2, cyt b*
AUS5	Aboriginal Australian	MOW19	*ND2, cyt b*
AUS6	Aboriginal Australian	MOW41	*cyt b*

4.3.3 Intraspecific Variation

The total numbers, the numbers of replacement (nonsynonymous), and the numbers of silent (synonymous) nucleotide differences per site between sequence pairs were estimated using a simple proportional distance (*p*-distance). This was obtained by dividing the number of nucleotide differences between two sequences by the total number of nucleotides compared. The number of nonsynonymous and synonymous sites for each gene were estimated using the method of Li (1993).

The amount of genetic variation within a species and its standard deviation were estimated from the number of polymorphic sites using Equations 2.1 and 2.2 as outlined in Section 2.6.3.1, and from the proportion of nucleotide differences between each pair of sequences using Equations 2.3 and 2.4 as outlined in Section 2.6.3.2.

4.3.4 Network Analysis

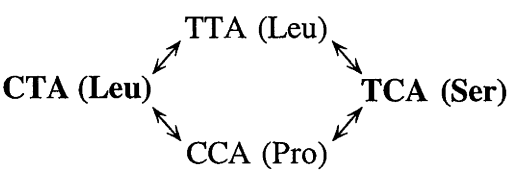
Networks are generated by partitioning the groups of sequence types character by character (Bandelt *et al.* 1995). An unmodified network contains all most parsimonious solutions and displays graphically the full information content of the sequence data. This approach highlights any incompatibility between pairs of characters, which enables identification of homoplasy (parallel mutation events or reversals) and can assist in identifying sequencing errors. This ability to retain alternative routes makes the networks more useful and less arbitrary, than forcing the data into a tree.

Theoretically, networks require as many dimensions as there are variable sites in the data, in order to represent the relationship among the sequences accurately. It is rare that so many dimensions are actually needed in practice, but the number may be substantial. If there are more than four dimensions, it is not only hard to illustrate the network, it is also difficult for the end user to interpret the diagram and gain a reasonable idea of the relationships illustrated (Bandelt *et al.* 1995). In practice networks should contain structures no more complex than three-dimensional. Resolution can be achieved by employing a reduction strategy based on compatibility and frequency arguments (Bandelt *et al.* 1995).

This approach was used to portray the human and chimpanzee mitochondrial *ND2* and *cyt b* sequence relationships.

4.3.5 Tests of Neutrality

The data were tested for departures from the neutral expectation that the ratio of nonsynonymous to synonymous polymorphisms within species should equal the ratio of nonsynonymous to synonymous fixed differences between species (McDonald and Kreitman 1991). Within and between species, nucleotide differences were counted as the number of mutational events occurring along the various branches of a network connecting humans, chimpanzees, and gorilla. All of these nucleotide differences were classified as either nonsynonymous or synonymous. When there is only one nucleotide difference, we can immediately decide whether the difference is nonsynonymous or synonymous. For example, if the codons compared are GTT (Val) and GTA (Val), there is one synonymous difference (Nei 1987). When there are two nucleotide differences between the codons compared, there are two possible ways to obtain the differences. At codon 110 of the *cyt b* gene there are two fixed differences between chimpanzees (CTA) and humans (TCA). Also, at codon 323 of the *cyt b* gene there are two fixed differences between the gorilla (CTA) and humans (TCA). The two possible pathways to reach from one codon to the other are:



We assume that the two pathways occur with equal probability. Thus, the first base of codons 110 and 323 (positions 15074 and 15713, respectively) may involve either a synonymous substitution or a nonsynonymous substitution.

In addition to the standard categories of nonsynonymous *vs.* synonymous mutations, further categories were defined based on the secondary structure of each protein. According to the predicted structure of the human ND2 protein there are 10 transmembrane domains and 11 surface domains (Persson and Argos 1994). Hence, there are total of four mutational categories: transmembrane nonsynonymous, transmembrane synonymous, surface nonsynonymous, and surface synonymous. The cytochrome *b* model consists of five internal, eight transmembrane, and four external domains (Esposti *et al.* 1993), resulting in six mutational categories.

Contingency tables were constructed in which one dimension consists of the structural mutational categories and the other dimension consists of the "fixed" *vs.* "polymorphic" categories. The intraspecific "polymorphic" class was further split into those mutations falling on "external" (*i.e.*, tip) branches *vs.* "internal" (*i.e.*, interior) branches (Castelloe and Templeton 1994; Templeton 1996). A tip haplotype is connected to only one other haplotype in the tree. An interior haplotype is connected to two or more other haplotypes in the tree, and hence represents an interior node in a topological sense (Templeton 1996). Two-by-two contingency tables were analysed using Fisher's exact test (FET) (Sokal and Rohlf 1995). Larger tables were analysed with an exact permutational test using the algorithm of Zaykin and Pudovkin (1993) and using 1000 random permutations of the data to simulate the null hypothesis of homogeneity. Uncorrected values were used in the statistical tests. This ensures that all observations are independent and results in conservative tests when there is an excess of polymorphic nonsynonymous differences since the number of fixed synonymous differences may be underestimated (Maynard-Smith 1994). The neutrality tests of Tajima (1989*a*) and Fu and Li (1993) were also performed for all nucleotide sites, nonsynonymous sites and synonymous sites in humans and chimpanzees as outlined in Section 2.6.4.

4.4 RESULTS

4.4.1 Intraspecific Variation

4.4.1.1 ND2

Nucleotide sequences were compared with previously published human (Anderson *et al.* 1981), chimpanzee, bonobo and gorilla (Horai *et al.* 1995) sequences. All polymorphic and fixed nucleotide sites are shown in Figure 4.1. Polymorphism data within humans and chimpanzees are summarised in Table 4.2.

Within humans there were 17 sequence types among 21 individuals, and two sequence types were shared among two or four individuals. Most differences between the sequences (92.7%) result from transition-type mutations, which is consistent with the general patterns of mtDNA sequence variation in humans (*e.g.*, Aquadro and Greenberg 1983; Greenberg *et al.* 1983; Vigilant *et al.* 1989; Horai and Hayasaka 1990; Horai *et al.* 1993; Watson *et al.* 1996). The uncorrected nucleotide diversity is: $\pi_T = 0.24\% \pm 0.15\%$ per site, $\pi_N = 0.17\% \pm 0.12\%$ per nonsynonymous site, and $\pi_S = 0.46\% \pm 0.34\%$ per synonymous site (Table 4.2).

Within chimpanzees there were 16 sequence types among 21 individuals, and four sequence types were shared among two or three individuals. Again, differences between the sequences (99.1%) result almost exclusively from transition-type mutations. The bias toward transitions has been noted in previous sequence comparisons of mtDNA in chimpanzees (see Section 3.4.1; Morin *et al.* 1994; Wise *et al.* 1997). The uncorrected nucleotide diversity (per site) is: $\pi_T = 1.02\% \pm 0.54\%$, $\pi_N = 0.31\% \pm 0.20\%$, and $\pi_S = 3.12\% \pm 1.69\%$ (Table 4.2).

4.4.1.2 Cyt b

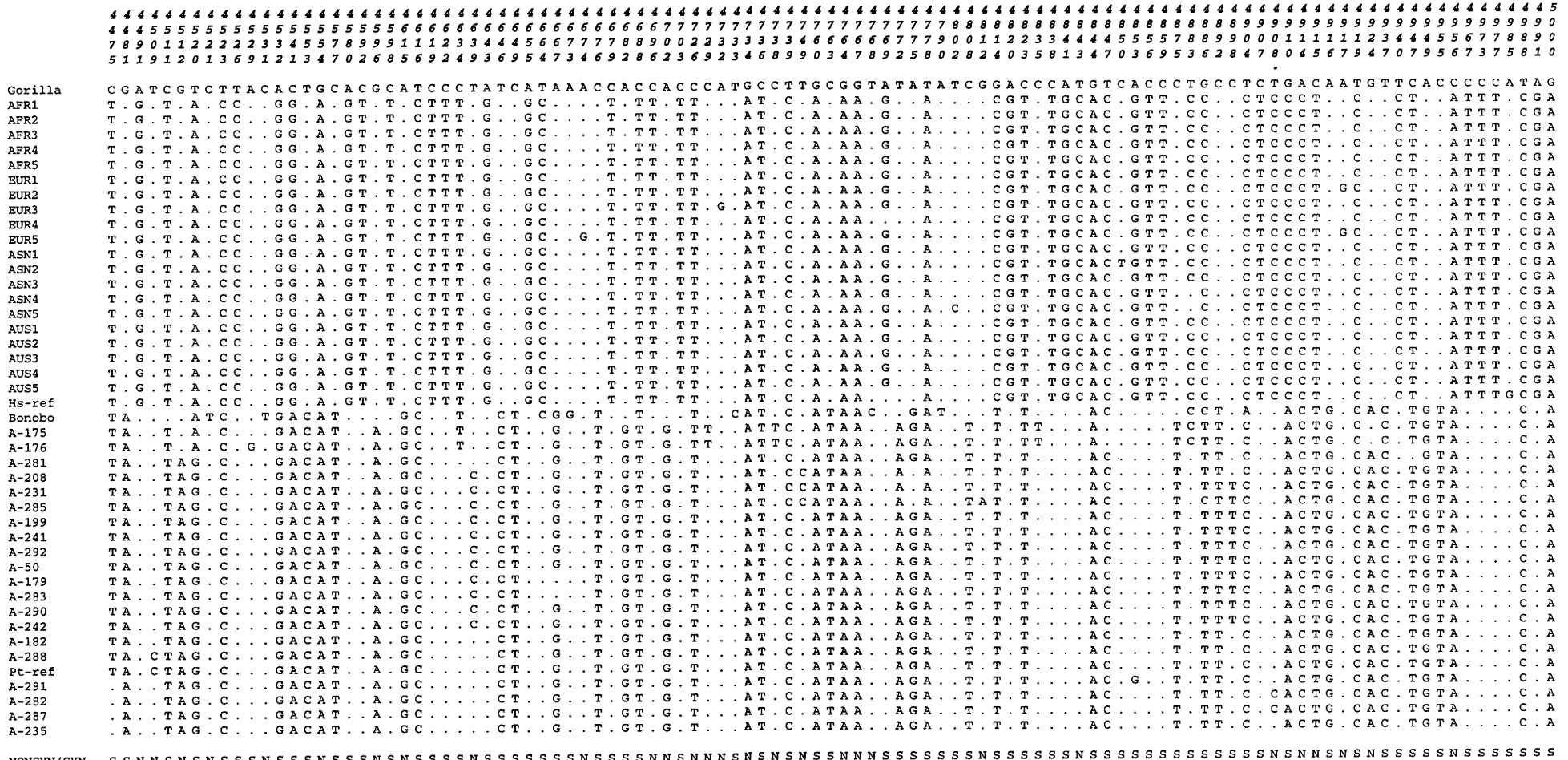
Nucleotide sequences were again compared with previously published human (Anderson *et al.* 1981), chimpanzee, bonobo and gorilla (Horai *et al.* 1995) sequences. All polymorphic and fixed nucleotide sites are shown in Figure 4.2. Polymorphism data within humans and chimpanzees are summarised in Table 4.2.

Within humans there were 17 sequence types among 21 individuals, and one sequence type was shared among five individuals. As above, most differences between the sequences (93.6%) result from transition-type mutations. The uncorrected nucleotide diversity (per site) is: $\pi_T = 0.35\% \pm 0.20\%$, $\pi_N = 0.11\% \pm 0.09\%$, and $\pi_S = 1.08\% \pm 0.66\%$ (Table 4.2).

Within chimpanzees there were 13 sequence types among 21 individuals, and four sequence types were shared among two or six individuals. Almost all of the polymorphisms were transitions (96.4%). The uncorrected nucleotide diversity (per site) is: $\pi_T = 1.10\% \pm 0.58\%$, $\pi_N = 0.17\% \pm 0.12\%$, and $\pi_S = 3.87\% \pm 2.05\%$ (Table 4.2).

Overall chimpanzees are more variable than humans. This high level of nucleotide diversity in chimpanzees derives partly from the presence of a few very divergent sequence types (A-175 and A-176) which differ at 28 out of 1041 sites (2.69%) for *ND2* and 37 out of 1140 sites (3.25%) for *cyt b*. This is considerably greater than the most divergent human sequence types, which differ at six out of 1041 sites (0.58%) for *ND2* and eight out of 1140 sites (0.7%) for *cyt b*. However, it is similar to the level of divergence reported for a small section of the mitochondrial *cyt b* gene (2.8%) (Morin *et al.* 1994) and the *ND3* gene (2.03%) (Nachman *et al.* 1996) between *P. t. verus* and either *P. t. troglodytes* or *P. t. schweinfurthii*. Divergence between *P. t. troglodytes* and *P. t. schweinfurthii* at *cyt b* is < 0.5% (Morin *et al.* 1994). It is therefore likely that the present sample includes *P. t. verus* and at least one of the two closely related subspecies, as noted previously (see Section 3.4.1; Wise *et al.* 1997).

To ensure that the chimpanzee sample represents a single interbreeding group, individuals A-175 and A-176 were excluded from all analyses. The estimates of nucleotide diversity within the *ND2* and *cyt b* genes of the west African subspecies *P. t. verus* are shown in Table 4.2. The nonsynonymous diversity estimates are similar to humans, whereas synonymous estimates are approximately 5.1-fold (*ND2*) and 2.4-fold (*cyt b*) higher than in humans.


Gorilla CGATCGTCTTACACTGCACGCATCCCTATCATAAACCACCACCCATGCCTTGCGGTATATATCGGACCCCATGTCACCCCTGCCCTCTGACAATGTTCACCCCCCATAG
AFR1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AFR2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AFR3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AFR4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AFR5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
EUR1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
EUR2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT.GC..CT..ATTT.CGA
EUR3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT.G.AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
EUR4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.....A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
EUR5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC..G.T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT.GC..CT..ATTT.CGA
ASN1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
ASN2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCACTGTT.CC..CTCCCT..C..CT..ATTT.CGA
ASN3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
ASN4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT..C..CTCCCT..C..CT..ATTT.CGA
ASN5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A.C..CGT.TGCAC.GTT..C..CTCCCT..C..CT..ATTT.CGA
AUS1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AUS2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AUS3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AUS4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
AUS5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
Hs-ref T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.....A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTTGCGA
Bonobo TA....ATC..TGACAT....GC..T..CT.CGG.T..T...T..CAT.C.ATAAC..GAT...T.T....AC.....CCT.A..ACTG.CAC.TGTA....C.A
A-175 TA..T.A.C...GACAT..A.GC..T..CT..G..T.GT.G.TT..ATTC.ATAA..AGA..T.T.TT...A.....TCTT.C..ACTG.C.C.TGTA....C.A
A-176 TA..T.A.C.G.GACAT..A.GC..T..CT..G..T.GT.G.TT..ATTC.ATAA..AGA..T.T.TT...A.....TCTT.C..ACTG.C.C.TGTA....C.A
A-281 TA..TAG.C...GACAT..A.GC....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC..GTA....C.A
A-208 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.CCATAA..A.A..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
A-231 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.CCATAA..A.A..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-285 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.CCATAA..A.A..TAT.T....AC....T.CTTC..ACTG.CAC.TGTA....C.A
A-199 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-241 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-292 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-50 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-179 TA..TAG.C...GACAT..A.GC...C.CT.....T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-283 TA..TAG.C...GACAT..A.GC...C.CT.....T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-290 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-242 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
A-182 TA..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
A-288 TA.CTAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
Pt-ref TA.CTAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
A-291 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC.G..T.TT.C..ACTG.CAC.TGTA....C.A
A-282 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C.CACTG.CAC.TGTA....C.A
A-287 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
A-235 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
NONSYN/SYN S S N N S N S N S S S N S S S N S S S N S N S S S S N S S S S S S S S N S S S S S N N N S N S N S S N N N S S S S S S S N S S S S S S N S S S S S S S S S S S S S S S N S N N S N S N S S S S S N S S S S S S S N N S S

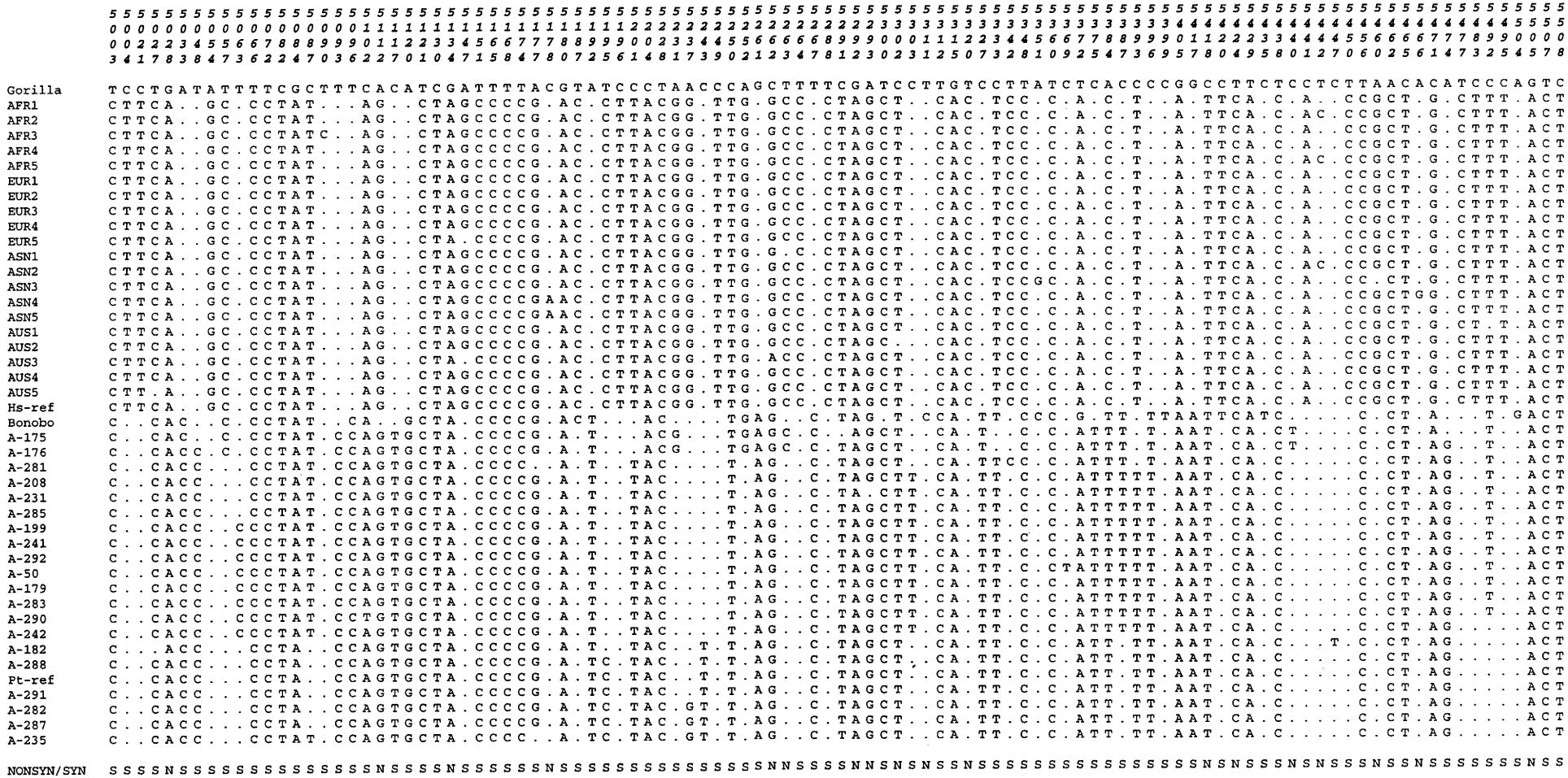
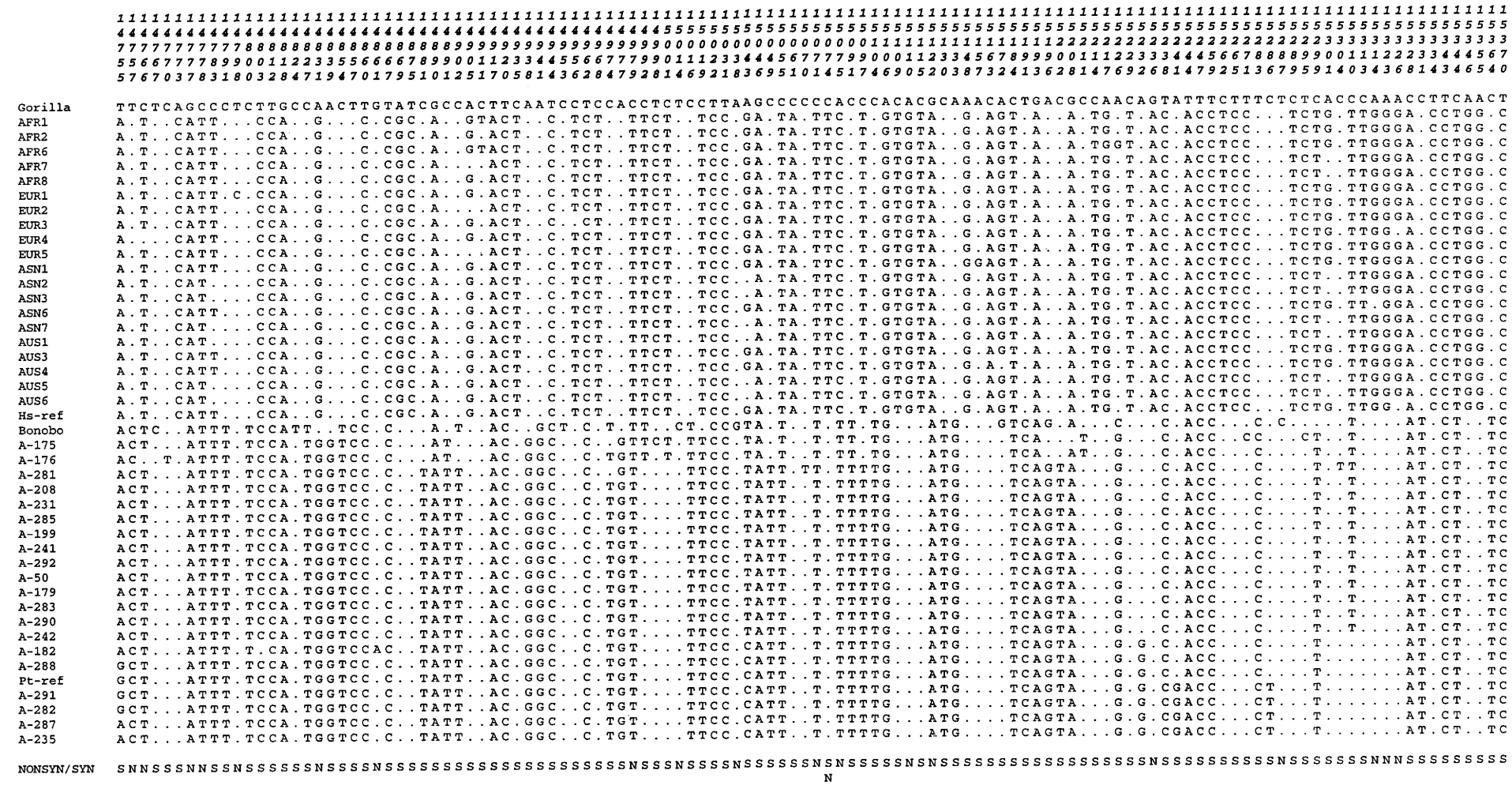


FIGURE 4.1 Variable nucleotide sites in the coding region of the *ND2* gene from humans, chimpanzees, bonobo (Horai *et al.* 1995) and gorilla (Horai *et al.* 1995). Nucleotide positions are numbered according to Anderson *et al.* (1981). A dot (.) indicates identity to the gorilla sequence, and differences are classified as either nonsynonymous (N) or synonymous (S). Both nonsynonymous and synonymous changes occur at positions 4886 and 4904. Human sequences are from Africa (AFR), Europe (EUR), Asia (ASN) and Australia (AUS). Hs-ref is a previously published human sequence (Anderson *et al.* 1981). The majority of the chimpanzee sequences (prefix "A") are from the west African subspecies *P. t. verus*, with the exception of A-175 and A-176 which are from *P. t. troglodytes* and/or *P. t. schweinfurthii* (see Section 3.4.1; Wise *et al.* 1997). Pt-ref is a previously published chimpanzee sequence (Horai *et al.* 1995).


Gorilla
AFR1
AFR2
AFR6
AFR7
AFR8
EUR1
EUR2
EUR3
EUR4
EUR5
ASN1
ASN2
ASN3
ASN6
ASN7
AUS1
AUS3
AUS4
AUS5
AUS6
Hs-ref
Bonobo
A-175
A-176
A-281
A-208
A-231
A-285
A-199
A-241
A-292
A-50
A-179
A-283
A-290
A-242
A-182
A-288
Pt-ref
A-291
A-282
A-287
A-235
NONSYN/SYN

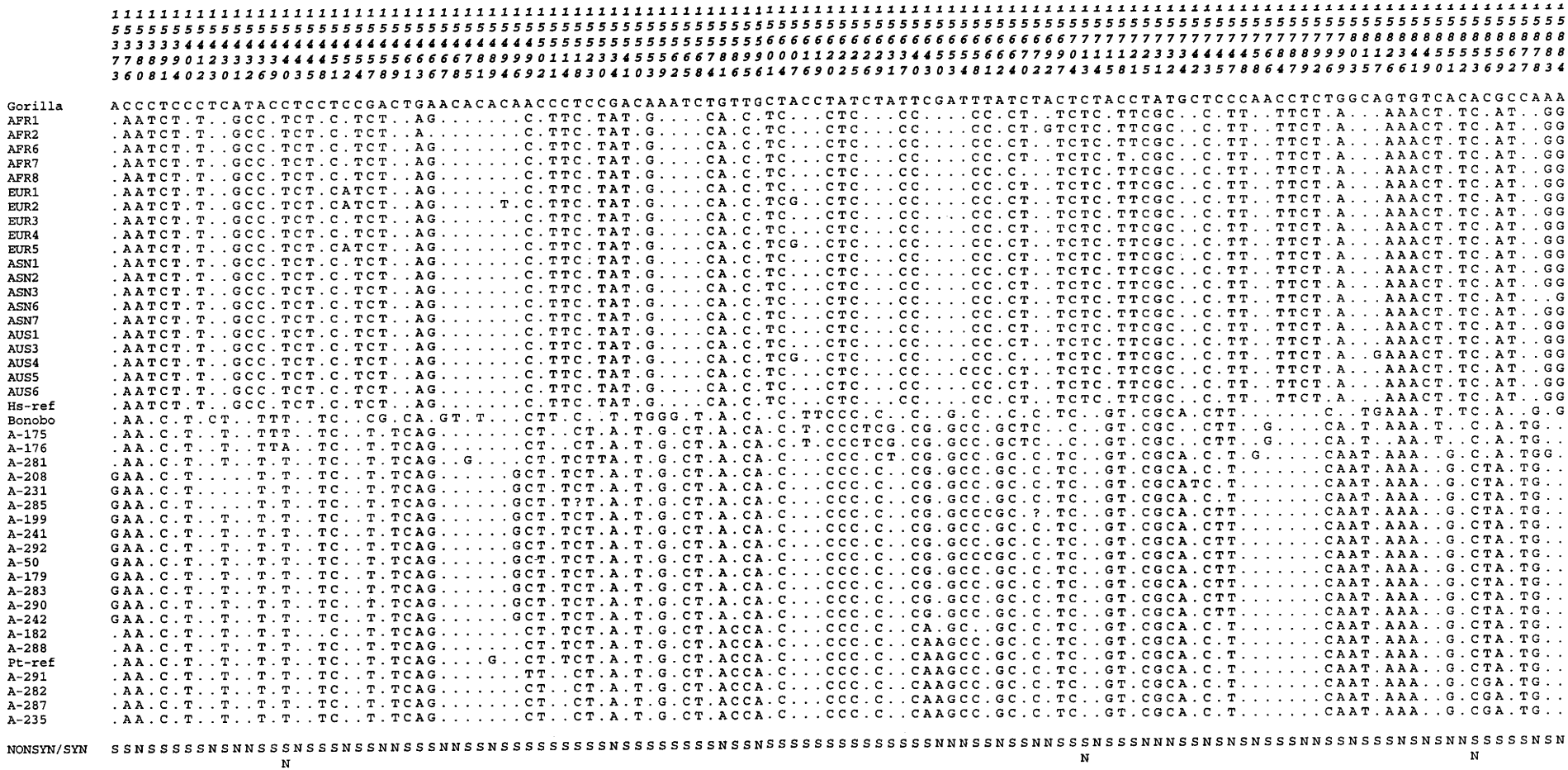


FIGURE 4.2 Variable nucleotide sites in the coding region of the *cyt b* gene from humans, chimpanzees, bonobo (Horai *et al.* 1995) and gorilla (Horai *et al.* 1995). Nucleotide positions are numbered according to Anderson *et al.* (1981). A dot (.) indicates identity to the gorilla sequence, and differences are classified as either nonsynonymous (N) or synonymous (S). Positions 15074, 15713 and 15853 involve either a nonsynonymous or synonymous change (see Section 4.3.5). Both types of changes occur at position 15440. Human sequences are from Africa (AFR), Europe (EUR), Asia (ASN) and Australia (AUS). Hs-ref is a previously published human sequence (Anderson *et al.* 1981). The majority of the chimpanzee sequences (prefix "A") are from the west African subspecies *P. t. verus*, with the exception of A-175 and A-176 which are from *P. t. troglodytes* and/or *P. t. schweinfurthii* (see Section 3.4.1; Wise *et al.* 1997). Pt-ref is a previously published chimpanzee sequence (Horai *et al.* 1995).

TABLE 4.2 Summary of *ND2* and *cyt b* variation in humans and chimpanzees.

	ND2[a]						*cyt b*[b]					
	S	η	η_e	$\pi \pm$ s.d.	$\theta \pm$ s.d.	max. π	S	η	η_e	$\pi \pm$ s.d.	$\theta \pm$ s.d.	max. π
Humans (n = 21)												
All sites	20	21	15	0.0024 ± 0.0015	0.0053 ± 0.0021	0.0058	23	26	18	0.0035 ± 0.0020	0.0056 ± 0.0022	0.0070
Nonsynonymous	10	10	7	0.0017 ± 0.0012	0.0035 ± 0.0016	0.0052	7	7	5	0.0011 ± 0.0009	0.0022 ± 0.0011	0.0047
Synonymous	10	11	8	0.0046 ± 0.0034	0.0109 ± 0.0049	0.0151	16	19	13	0.0108 ± 0.0066	0.0163 ± 0.0066	0.0245
All Chimpanzees (n = 21)												
All sites	48	52	21	0.0102 ± 0.0054	0.0128 ± 0.0046	0.0269	60	64	26	0.0110 ± 0.0058	0.0146 ± 0.0052	0.0325
Nonsynonymous	12	12	4	0.0031 ± 0.0020	0.0042 ± 0.0018	0.0116	9	9	5	0.0017 ± 0.0012	0.0029 ± 0.0013	0.0082
Synonymous	36	40	17	0.0312 ± 0.0169	0.0394 ± 0.0145	0.0796	51	55	21	0.0387 ± 0.0205	0.0519 ± 0.0185	0.1154
***P. t. verus* (n = 19[c])**												
All sites	28	39	17	0.0073 ± 0.0040	0.0077 ± 0.0030	0.0144	32	45	17	0.0073 ± 0.0040	0.0080 ± 0.0030	0.0175
Nonsynonymous	6	7	4	0.0018 ± 0.0013	0.0022 ± 0.0011	0.0052	4	6	3	0.0009 ± 0.0008	0.0013 ± 0.0008	0.0035
Synonymous	22	32	13	0.0235 ± 0.0131	0.0248 ± 0.0098	0.0417	28	39	14	0.0262 ± 0.0144	0.0294 ± 0.0112	0.0628

NOTE.—S is the number of polymorphic sites; η is the total number of mutations and η_e is the number of unambiguous mutations in the external branches (inferred from Figures 4.3 and 4.4); π is the average number of pairwise nucleotide differences per site; θ is the average number of nucleotides segregating per site; s.d. is the standard deviation; and max. π is the maximum number of pairwise nucleotide differences per site. Values of π and θ, and their standard deviations were calculated as outlined in Section 4.3.3. No multiple-hit correction was made.

[a] The total number of sites compared for the *ND2* gene was 1041. The numbers of nonsynonymous (≈787) and synonymous (≈254) sites were estimated using the method of Li (1993) .

[b] The total number of sites compared for the *cyt b* gene was 1140. The numbers of nonsynonymous (≈867) and synonymous (≈273) sites were estimated using the method of Li (1993).

[c] Excluding individuals A-175 and A-176, as they are not from the *P. t. verus* subspecies.

4.4.2 Network Analysis

Before the contingency analyses can be performed it is necessary to resolve any ambiguities in the networks as this will determine the topological categories into which sequence differences are sorted.

4.4.2.1 ND2

Figure 4.3 *A* shows the unrooted network for 17 human *ND2* sequence types. There is a box of ambiguity involving nucleotide positions 4917 and 5147. Since position 4917 is a nonsynonymous change it is more parsimonious to assume that two mutational events have occurred at position 5147. This could involve either a reversal or two parallel mutations, which will be referred to as resolutions Ia and IIa, respectively. Figure 4.3 *B* shows the unrooted network for 16 chimpanzee *ND2* sequence types. Homoplasies occur at nucleotide position 5177 (two parallel mutations marked by asterisks), and probably positions 5087 (reversal) and 5492 (two parallel mutations). The box of ambiguity containing position 4898 can be broken in two equally parsimonious ways involving either a reversal or two parallel mutations, which will be referred to as resolutions Ib and IIb, respectively. These ambiguities are important because the thick lines in the boxes can either be an interior branch or part of a tip branch, depending on which resolution is chosen. In all analyses, whenever these ambiguities are relevant, tests are performed under all resolutions to ensure robustness of the test results to this network uncertainty.

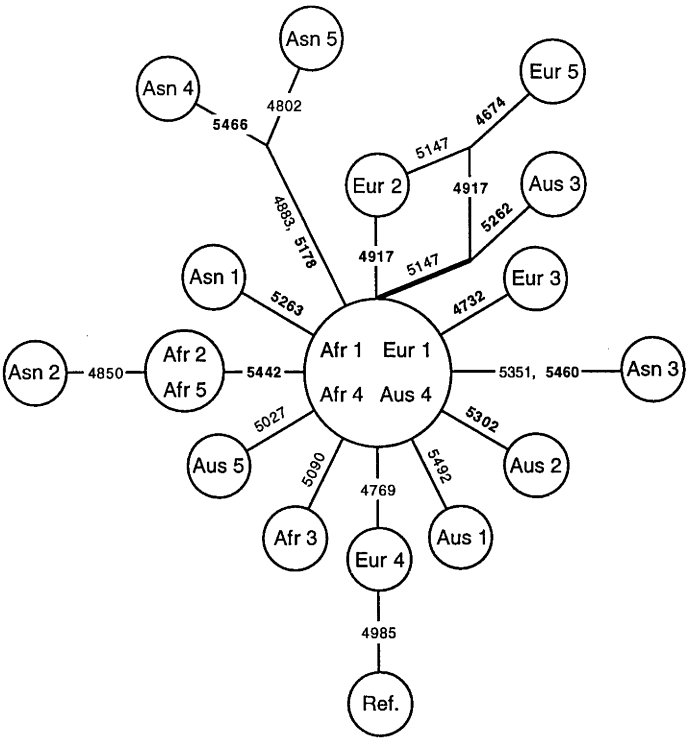
Figure 4.3 *C* shows a reduced *ND2* network containing the bonobo and gorilla sequences found in Horai *et al.* (1995). There are 11 fixed nonsynonymous nucleotide differences between humans and chimpanzees. The number of fixed synonymous differences varies between 84 and 91 depending on the branch placement of some mutations. In the contingency analyses to follow, the minimum number of synonymous substitutions is used to ensure that the tests remain conservative.

4.4.2.2 Cyt b

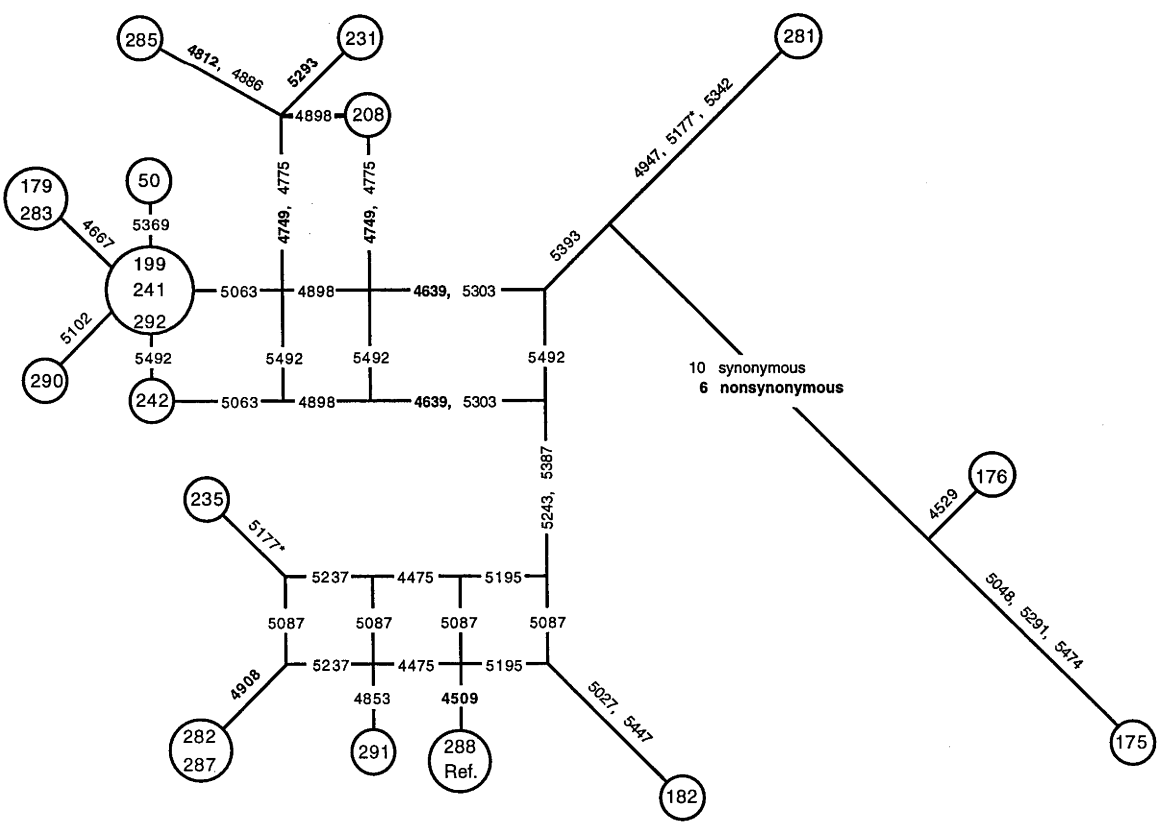
Figure 4.4 *A* shows the unrooted network for 17 human *cyt b* sequence types. Homoplasies occur at nucleotide positions 15607 and 15670 (parallel mutations marked by asterisks), and probably position 14905 (two parallel mutations). Figure 4.4 *B* shows the unrooted network for 13 chimpanzee *cyt b* sequence types. Homoplasies occur at nucleotide positions 14974 (two parallel mutations), 15514 (reversal) and 15745 (reversal) (marked by asterisks), and probably position 14755b (reversal).

Figure 4.4 *C* shows a reduced *cyt b* network containing the bonobo and gorilla sequences found in Horai *et al.* (1995). The number of fixed nonsynonymous differences between humans and chimpanzees varies between 25 and 30, and the number of synonymous substitutions varies between 96 and 110 depending on the classification of positions 15074 and 15713 (see Section 4.3.5) and on the branch placement of some mutations. Contingency tests are performed using both the minimum and maximum counts to ensure robustness of the test results to this ambiguity.

A


Asn 5
Asn 4
4802
5466
Eur 5
4674
5147
Eur 2
4917
Aus 3
5262
4883, 5178
4917
5147
Asn 1
5263
Eur 3
4732
Asn 2
4850
Afr 2
Afr 5
5442
Afr 1 Eur 1
Afr 4 Aus 4
5351, 5460
Asn 3
5027
5302
Aus 5
5090
4769
5492
Aus 2
Afr 3
Eur 4
Aus 1
4985
Ref.

B


285
4812, 4886
5293
231
4898
208
4947, 5177*, 5342
281
179
283
50
5369
4775
4749,
4775
4749,
4667
199
241
292
5063
4898
4639, 5303
5393
5102
5492
5492
5492
5492
290
242
5063
4898
4639, 5303
10 synonymous
6 nonsynonymous
5387
5243,
176
4529
235
5177*
5237
4475
5195
5087
5087
5087
5087
5048, 5291, 5474
4908
5237
4475
5195
4853
4509
5027, 5447
282
287
291
288
Ref.
182
175

C

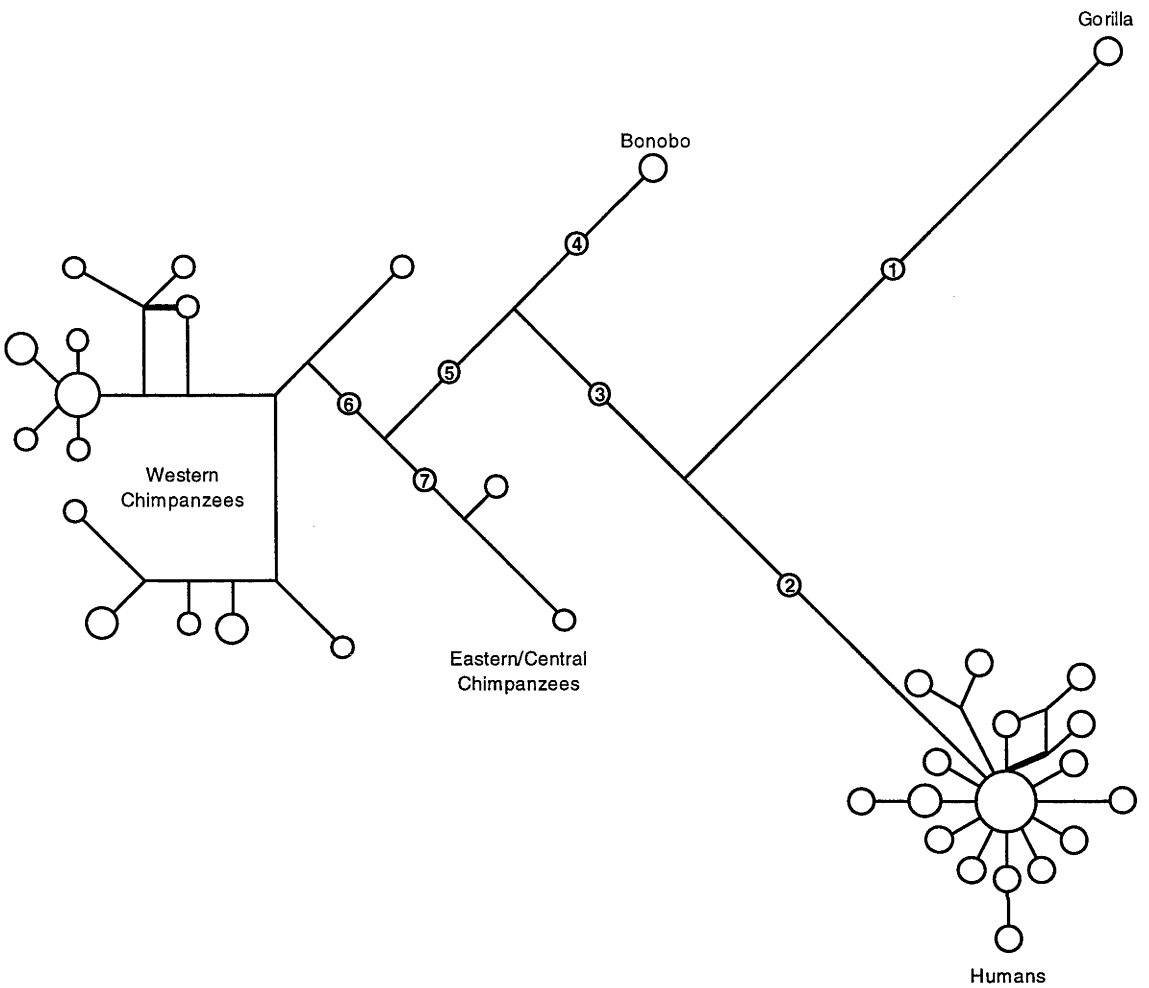
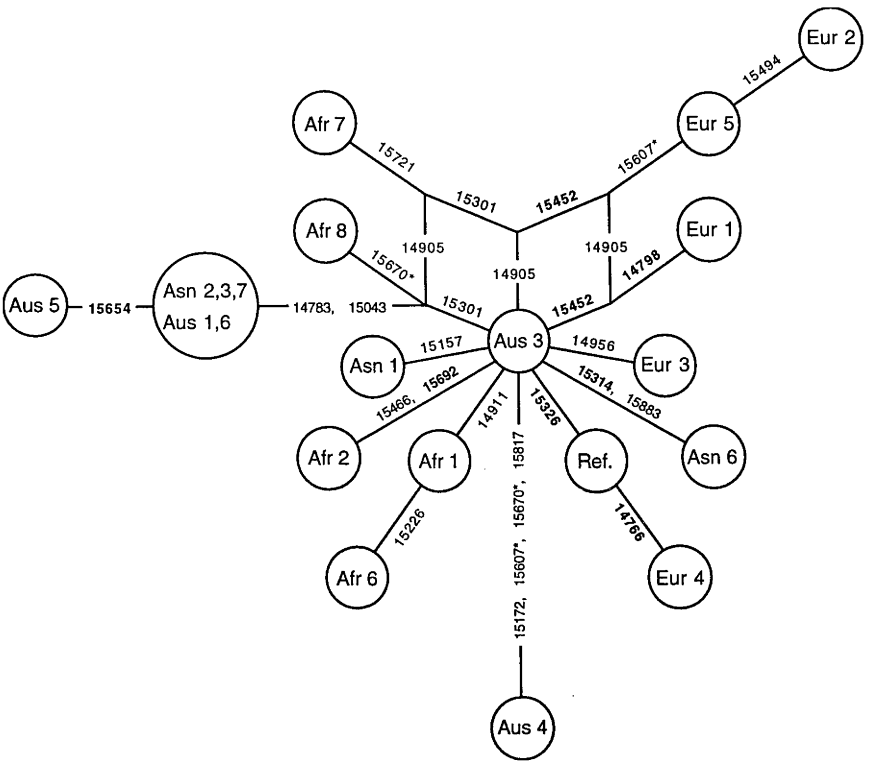
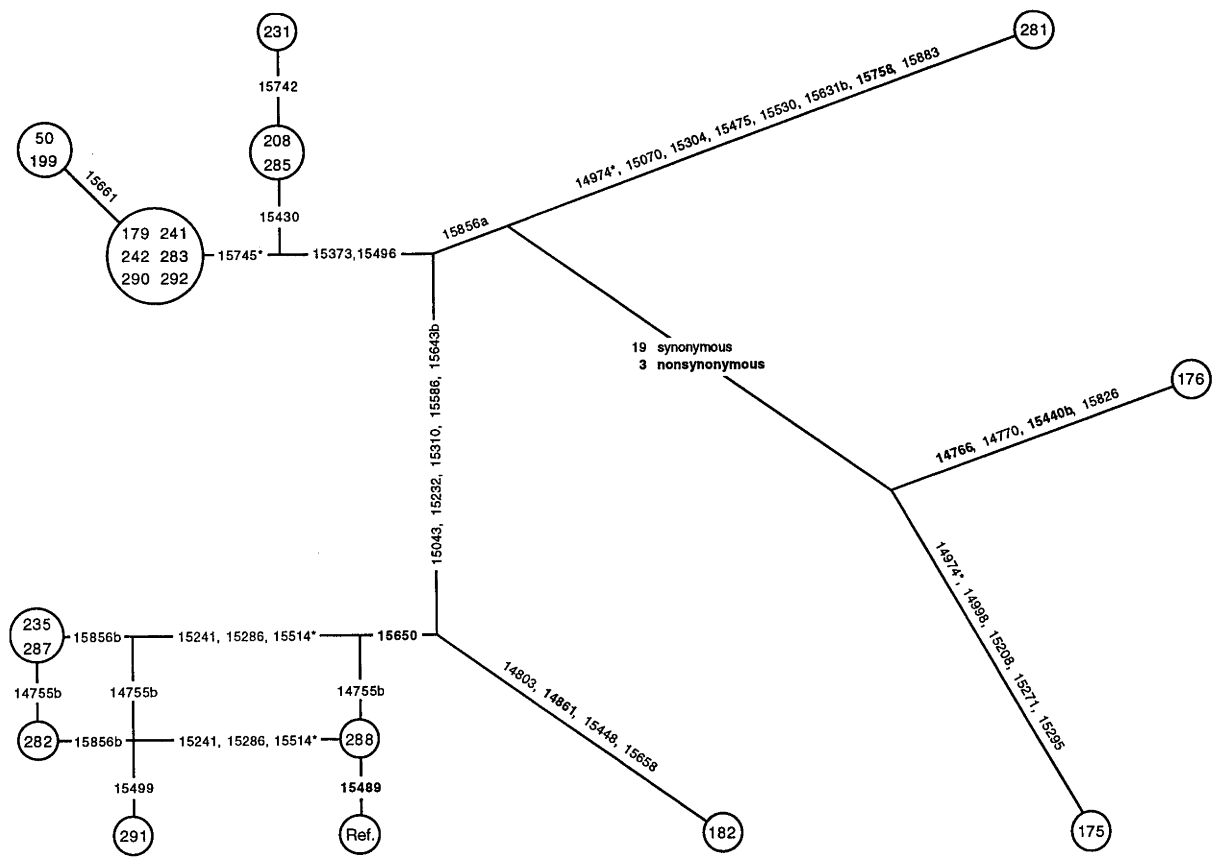


FIGURE 4.3 *A*, Unrooted median network of 17 human *ND2* sequence types. *B*, Unrooted median network of 16 chimpanzee *ND2* sequence types. *C*, Reduced human and chimpanzee *ND2* network containing bonobo and gorilla sequences (not to scale). Circles denote sequence types, and individuals are identified as in Figure 4.1. Differences between sequence types are numbered according to Anderson *et al.* (1981). Amino acid replacement (nonsynonymous) mutations are shown in bold, and likely parallel mutations or reversals are marked by asterisks. Ambiguities in the networks (see Section 4.4.2.1) are indicated by thick lines. Unambiguous nonsynonymous and synonymous nucleotide changes along branches ① to ⑦ are listed in Appendix D1. The following synonymous sites are ambiguous with respect to their branch placement: 4511–branches ① and ④, or ② and ⑤; 4664, 5187, 5420–branches ① and ⑤, or ② and ④; 4541, 4814, 4910, 5384, 5471–branches ① and ②, or ① and ③, or ② and ③.

A


Eur 2
15494
Afr 7
Eur 5
15721
15607*
15301
15452
Afr 8
Eur 1
14905
14905
14905
15670*
14798
Aus 5
15654
Asn 2,3,7
Aus 1,6
14783, 15043
15301
15452
Aus 3
Asn 1
15157
14956
Eur 3
15466, 15692
15314, 15883
14911
15326
15817
Afr 2
Afr 1
Ref.
Asn 6
15226
14766
Afr 6
Eur 4
15172, 15607*, 15670*,
Aus 4

B


231
15742
281
50
199
15661
208
285
15430
14974*, 15070, 15304, 15475, 15530, 15631b, 15758, 15883
179 241
242 283
290 292
15745*
15373,15496
15856a
19 synonymous
3 nonsynonymous
176
15043, 15232, 15310, 15586, 15643b
14766, 14770, 15440b, 15826
14974*, 14998, 15208, 15271, 15295
235
287
15856b
15241, 15286, 15514*
15650
14755b
14755b
14755b
282
15856b
15241, 15286, 15514*
288
14803, 14861, 15448, 15658
15499
15489
291
Ref.
182
175

C

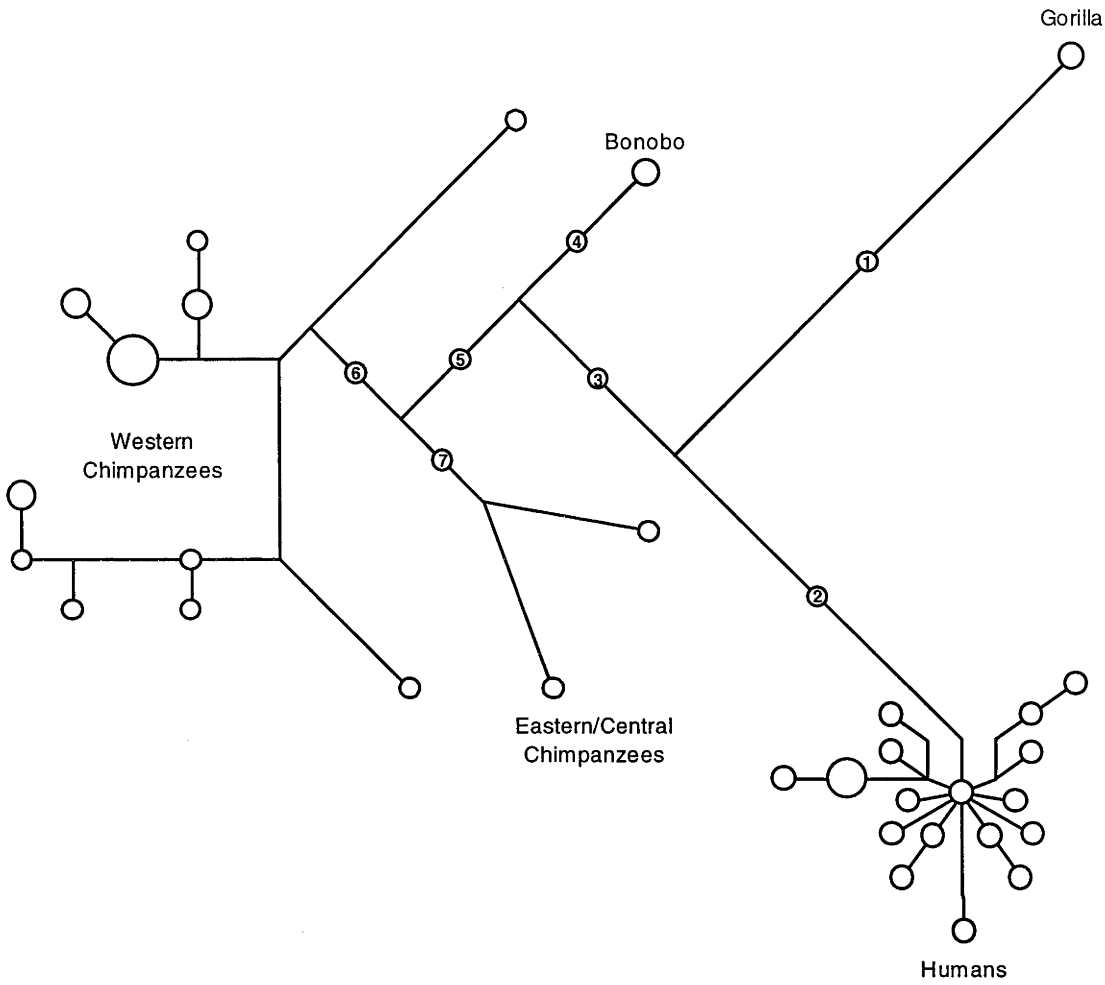


FIGURE 4.4 *A*, Unrooted median network of 17 human *cyt b* sequence types. *B*, Unrooted median network of 13 chimpanzee *cyt b* sequence types. *C*, Reduced human and chimpanzee *cyt b* network containing bonobo and gorilla sequences (not to scale). Circles denote sequence types, and individuals are identified as in Figure 4.2. Differences between sequence types are numbered according to Anderson *et al.* (1981). Amino acid replacement (nonsynonymous) mutations are shown in bold, and likely parallel mutations or reversals are marked by asterisks. Unambiguous nonsynonymous and synonymous nucleotide changes along branches ① to ⑦ are listed in Appendix D2. The following nonsynonymous sites are ambiguous with respect to their branch placement: 14831–branches ① and ④, or ② and ⑤; 15852, 15884–branches ① and ⑤, or ② and ④; 15075–branches ① and ②, or ① and ③, or ② and ③. The following synonymous sites are ambiguous with respect to their branch placement: 15019, 15299, 15466, 15799–branches ① and ④, or ② and ⑤; 15172, 15511, 15553–branches ① and ⑤, or ② and ④; 15226–branches ③ and ④, or ③ and ⑤; 15043, 15436,15622–branches ① and ②, or ① and ③, or ② and ③.

4.4.3 Contingency Tests of Neutrality

4.4.3.1 ND2

The contingency tables and test results are shown in Table 4.3 for the full contrast of all four mutational categories (transmembrane nonsynonymous, surface nonsynonymous, transmembrane synonymous, surface synonymous) *vs.* all three topological categories (tip, interior, fixed). The null hypothesis of homogeneity is rejected when polymorphism data from humans are compared with fixed differences between species, but not when chimpanzee polymorphisms are compared with interspecific differences.

TABLE 4.3 Contingency analysis of the full mutational *vs.* the full network topological categories within and between species at the *ND2* gene.

			Humans		Chimpanzees	
Mutation	**Region**	**Fixed**	**Tip**	**Interior**	**Tip**	**Interior**
Nonsynonymous	Transmembrane	10	7	3	4	2
	Surface	1	0	0	0	1
Synonymous	Transmembrane	56	4	1	10 (9)	14 (15)
	Surface	28	4 (5)	2 (1)	4	4
	Permutational probability:		0.049 (0.045)		0.577 (0.476)	

NOTE.—Exact permutational tests were used to test the null hypothesis of homogeneity (Zaykin and Pudovkin 1993). Categories which were affected by ambiguities in the networks (see Section 4.4.2.1) are indicated by a number followed by a second number in parenthesis. The first number is the count under resolution Ia for humans and resolution Ib for chimpanzees, and the number in parenthesis is the count under resolutions IIa and IIb. The probability values are presented in the same manner, with the first probability referring to resolution I, and the probability in parenthesis to resolution II.

The more standard McDonald and Kreitman (1991) test collapses the transmembrane and surface categories into the nonsynonymous/synonymous mutational categories and collapses the tip and interior categories into a single polymorphic category, yielding the 2×2 table given in Table 4.4. The tests reveal a significantly higher ratio of nonsynonymous to synonymous nucleotide differences within humans than is seen between species, but not within chimpanzees.

TABLE 4.4 Contingency analysis of the nonsynonymous/synonymous mutations *vs.* fixed/polymorphic topological categories within and between species at the *ND2* gene.

	Fixed (Between)	**Polymorphic (Within)**	
	Human and Chimpanzee	**Humans**	**Chimpanzees**
Nonsynonymous	11	10	7
Synonymous	84	11	32
	FET probability:	0.0005	0.4033

NOTE.—Fisher's exact tests (FET) were used to test the null hypothesis that the ratio of nonsynonymous to synonymous nucleotide differences is the same within and between species.

The contingency test for the full model can be subdivided to investigate the evolutionary dynamics of the different mutational categories (Templeton 1987, 1996). First, the impact of the structural region of the molecule upon the evolutionary dynamics of nonsynonymous and synonymous mutations can be examined by a contingency test of the first and second rows of Table 4.3 (which contrasts the evolutionary dynamics of nonsynonymous mutations in the transmembrane *vs.* surface regions) and a separate contingency analysis of the third and fourth rows (synonymous mutations across the two structural regions). The permutational probability values for the contingency analysis of nonsynonymous mutations in the transmembrane *vs.* surface regions are 1.000 and 0.369 for the human and chimpanzee data, respectively. For synonymous mutations across structural regions the permutational probabilities are 0.379 and 0.456 for human network resolutions Ia and IIa, respectively, and 0.673 and 0.602 for chimpanzee network resolu-

tions Ib and IIb, respectively. None of these results is significant at the 5% level, and this may be due in part to the small numbers of observations in some categories in the contingency table. To enhance statistical power, the data was further pooled into polymorphic (tip + interior) *vs.* fixed, and young (tip) *vs.* old (interior + fixed) categories. None of the tests was significant.

A second nested series of contingency tests examines the evolutionary dynamics of nonsynonymous *vs.* synonymous mutations within the transmembrane regions (rows one and three of Table 4.3) and within the surface regions (rows two and four of Table 4.3) (Templeton 1987, 1996). The transmembrane results are highly significant for the human data (Permutational probability = 0.000), but neither network resolution yields significant results for the chimpanzee data (Permutational probabilities = 0.411 and 0.372 for resolutions Ib and IIb, respectively). None of the surface results is significant (Permutational probability values are 1.000 for both human network resolutions, and 0.429 for the chimpanzee data). The pooling categories of polymorphic/fixed and young/old were also only significant for the human transmembrane data (FET probabilities = 0.0001 and 0.0037, respectively).

4.4.3.2 Cyt b

The contingency tables and test results are shown in Table 4.5 for the full contrast of all six mutational categories (synonymous and nonsynonymous mutations in the transmembrane, internal and external surface domains) *vs.* all three topological categories (tip, interior, fixed). Applied to the current data, this approach does not reject the neutral equilibrium model for either species. One way of regaining power in a contingency framework is to pool categories. The more standard 2 × 2 McDonald and Kreitman (1991) test (see Table 4.6) and the mutational contrast of young *vs.* old (Templeton 1996) did not produce significant results.

TABLE 4.5 Contingency analysis of the full mutational *vs.* the full network topological categories within and between species at the *cyt b* gene.

			Humans		Chimpanzees	
Mutation	**Region**	**Fixed**	**Tip**	**Interior**	**Tip**	**Interior**
Nonsynonymous	Transmembrane	18 (17)	2	2	3	3
	Internal surface	8 (4)	3	0	0	0
	External surface	4 (4)	0	0	0	0
Synonymous	Transmembrane	51 (61)	7	4	7	16
	Internal surface	12 (13)	1	1	4	2
	External surface	33 (36)	5	1	3	7
	Permutational probability:		0.852 (0.395)		0.481 (0.664)	

NOTE.—Exact permutational tests were used to test the null hypothesis of homogeneity (Zaykin and Pudovkin 1993). Categories which were affected by ambiguities in the network (see Section 4.4.2.2) are indicated by a number followed by a second number in parenthesis. The first number is the count assuming the maximum number of nonsynonymous substitutions and the minimum number of synonymous substitutions, and the number in parenthesis is the opposite count. The probability values are presented in the same manner. The probabilities do not vary significantly when other combinations are used.

TABLE 4.6 Contingency analysis of the nonsynonymous/synonymous mutations *vs.* fixed/polymorphic topological categories within and between species at the *cyt b* gene.

	Fixed (Between)	**Polymorphic (Within)**	
	Human and Chimpanzee	**Humans**	**Chimpanzees**
Nonsynonymous	30 (25)	7	6
Synonymous	96 (110)	19	39
	FET probability:	0.8026 (0.4195)	0.2007 (0.5007)

NOTE.—Fisher's exact tests (FET) were used to test the null hypothesis that the ratio of nonsynonymous to synonymous nucleotide differences is the same within and between species.

4.4.4 The Tajima and the Fu and Li Tests of Neutrality

The Tajima (1989*a*) test examines whether the average number of pairwise nucleotide differences between sequences (π) is larger or smaller than expected from the observed number of polymorphic sites (θ). Under the assumption of a random mating population at equilibrium, the difference between π and θ (D_T) is expected to be zero. A positive value of D_T indicates possible balancing selection or population subdivision. A negative value suggests recent directional selection, a population bottleneck or background selection of slightly deleterious alleles (Tajima 1989*a*). The Fu and Li (1993) test takes a genealogical approach and is based on the principle of comparing the number of mutations on internal branches with those on external branches. Compared with a neutral model of evolution, directional selection would result in an excess of external mutations while balancing selection would result in an excess of internal mutations. Ideally, an outgroup is used so that the number of mutations in the external branches can be unambiguously determined. Since it is not clear which tests are most powerful, Tajima's (1989*a*) D_T test and the D_F and F tests proposed by Fu and Li (1993) were used to investigate departures from neutrality at the *ND2* gene and the *cyt b* gene in humans and chimpanzees (Table 4.7).

A significantly negative D_T, D_F, and F is observed for the human *ND2* data. This is consistent with a pattern of there being too many rare nucleotide polymorphisms with respect to predictions of the neutral theory (see *e.g.*, Braverman *et al.* 1995). An excess of low frequency polymorphisms is also observed for the *cyt b* data, although in some cases the tests are not significant. In the case of chimpanzees, none of the tests are significant and thus, by this criterion, the data are consistent with a neutral model of mtDNA evolution.

TABLE 4.7 D_T, D_F, and F for the human and chimpanzee *ND2* and *cyt b* genes.

	ND2			*cyt b*		
	D_T	D_F	F	D_T	D_F	F
Human ($n = 21$)						
All sites	−2.071 (s)	−2.858 (s)	−3.185 (s)	−1.423	−2.819 (s)	−2.980 (s)
Nonsynonymous	−1.855 (s)	−2.321 (s)	−2.605 (s)	−1.719	−2.154 (s)	−2.406 (s)
Synonymous	−1.950 (s)	−2.537 (s)	−2.884 (s)	−1.125	−2.609 (s)	−2.742 (s)
***P. t. verus* ($n = 19$[a])**						
All sites	−0.190	−1.043	−1.378	−0.376	−0.648	−1.125
Nonsynonymous	−0.590	−1.380	−1.516	−0.958	−0.984	−1.447
Synonymous	−0.045	−0.816	−1.187	−0.245	−0.507	−0.953

NOTE.—D_T, Tajima's (1989*a*) test statistic. D_F and F, Fu and Li's (1993) test statistics. S, significant at the 5% level.

[a] Individuals A-175 and A-176 were excluded from the analyses, as they are not from the *P. t. verus* subspecies.

4.5 DISCUSSION

The contingency approach to testing neutrality depends on accurate and unbiased counts of the numbers of mutations in various categories (Templeton 1996). The network approach enables identification of any ambiguity in the tree topology which might affect this analysis. For the *ND2* data set, the human and chimpanzee networks each contained two alternatives that affected the numbers of mutations in some of the categories (see Section 4.4.2.1). For the *cyt b* data set, the number of interspecific differences provided a source of ambiguity (see Section 4.4.2.2). To assess robustness of conclusions to error in tree topology estimation, contingency analyses were repeated over all alternatives (Templeton 1996). In this case, the conclusions about the evolution of the *ND2* and *cyt b* genes in humans and chimpanzees are robust to this topological ambiguity.

The data presented here for the full contingency analysis provide a clear rejection of the null hypothesis that the human *ND2* gene is evolving according to a strictly neutral model of molecular evolution (see Table 4.3). This strong departure from neutrality is also seen in the McDonald and Kreitman (1991) test (see Table 4.4), and is consistent with an excess of nonsynonymous polymorphisms (or deficiency of synonymous polymorphisms) within humans compared with interspecific differences. This pattern is also seen in the *cyt b* gene (although a neutral model of molecular evolution is not rejected) (see Tables 4.5 and 4.6), and appears to be widespread in the human mitochondrial genome (Nachman *et al.* 1996).

These tests do not apply for highly diverged sequences: when fully saturated, the ratio of nonsynonymous to synonymous nucleotide differences is expected to be as much as two times greater in between-species than in within-species comparisons. This effect is not important when the species to be compared are closely related and the sequences are not close to saturation (Maynard-Smith 1994). Thus, an analysis of mutational saturation of nonsynonymous and synonymous substitutions should be done before interpreting a rejection of the null hypothesis as evidence for selection. Such an analysis of substitutions in the *ND2* gene is presented in Table 4.8 and Figure 4.5. Similarly, analysis of the *cyt b* gene is presented in Table 4.9 and Figure 4.6. In both cases, nonsynonymous substitutions are far from saturated between divergent species. Although synonymous substitutions do not appear to be saturated in the human × chimpanzee comparison, they are likely to be undercounted, resulting in conservative tests.

TABLE 4.8 Estimates of the number of nonsynonymous and synonymous nucleotide substitutions per site between five species of primates at the *ND2* gene.

	Human	Chimpanzee	Gorilla	Orangutan	Siamang
Human	—	0.308 ± 0.039	0.423 ± 0.049	0.527 ± 0.061	0.617 ± 0.066
Chimpanzee	0.025 ± 0.006	—	0.361 ± 0.043	0.473 ± 0.054	0.606 ± 0.068
Gorilla	0.044 ± 0.008	0.044 ± 0.008	—	0.474 ± 0.053	0.640 ± 0.068
Orangutan	0.103 ± 0.013	0.100 ± 0.013	0.109 ± 0.014	—	0.486 ± 0.053
Siamang	0.108 ± 0.013	0.115 ± 0.014	0.113 ± 0.014	0.141 ± 0.016	—

NOTE.—Values below and above the diagonal represent nonsynonymous and synonymous nucleotide substitutions, respectively, calculated using the method of Li (1993).

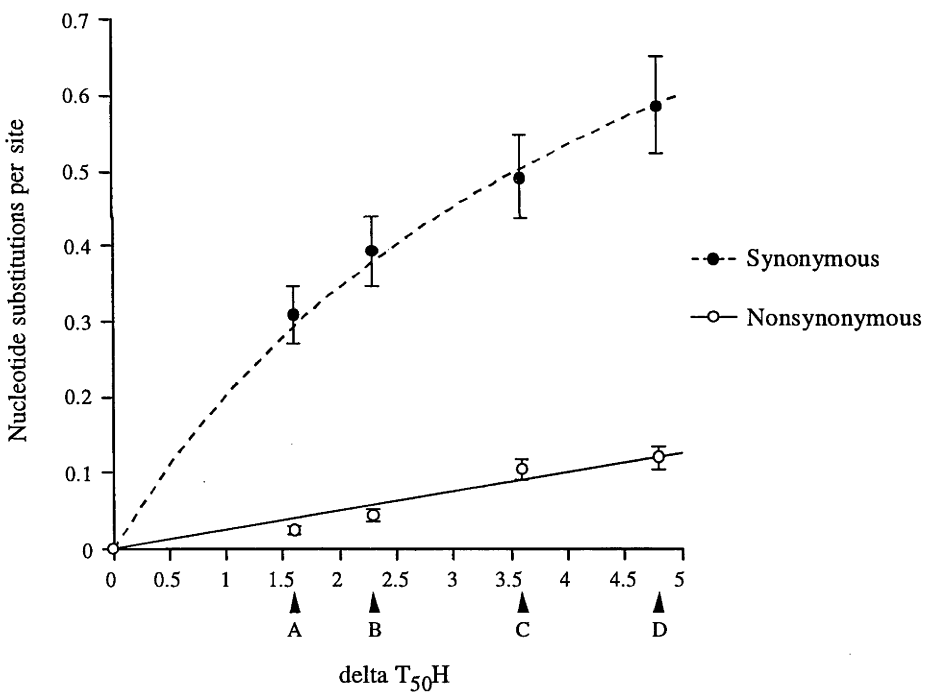


FIGURE 4.5 Accumulation of synonymous (broken line) and nonsynonymous (solid line) nucleotide substitutions in the *ND2* gene with delta $T_{50}H$ values derived from DNA-DNA hybridisation (Sibley and Ahlquist 1987). **A** human × chimpanzee comparison. **B** human + chimpanzee × gorilla comparisons. **C** human + chimpanzee + gorilla × orangutan comparisons. **D** human + chimpanzee + gorilla + orangutan × siamang comparisons.

TABLE 4.9 Estimates of the number of nonsynonymous and synonymous nucleotide substitutions per site between five species of primates at the *cyt b* gene.

	Human	Chimpanzee	Gorilla	Orangutan	Siamang
Human	—	0.318 ± 0.038	0.384 ± 0.044	0.438 ± 0.048	0.537 ± 0.056
Chimpanzee	0.048 ± 0.008	—	0.361 ± 0.041	0.454 ± 0.051	0.569 ± 0.058
Gorilla	0.053 ± 0.009	0.043 ± 0.008	—	0.454 ± 0.049	0.537 ± 0.056
Orangutan	0.083 ± 0.011	0.070 ± 0.010	0.068 ± 0.010	—	0.576 ± 0.060
Siamang	0.087 ± 0.012	0.077 ± 0.011	0.081 ± 0.011	0.103 ± 0.013	—

NOTE.—Values below and above the diagonal represent nonsynonymous and synonymous nucleotide substitutions, respectively, calculated using the method of Li (1993).

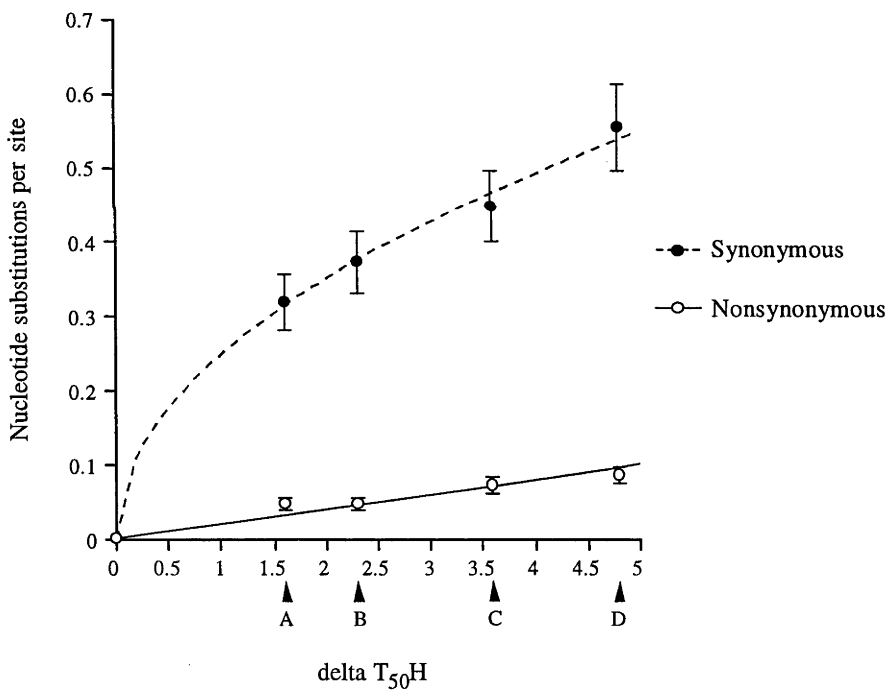


FIGURE 4.6 Accumulation of synonymous (broken line) and nonsynonymous (solid line) nucleotide substitutions in the *cyt b* gene with delta $T_{50}H$ values derived from DNA-DNA hybridisation (Sibley and Ahlquist 1987). **A** human × chimpanzee comparison. **B** human + chimpanzee × gorilla comparisons. **C** human + chimpanzee + gorilla × orangutan comparisons. **D** human + chimpanzee + gorilla + orangutan × siamang comparisons.

Further insights into the biological basis of the rejection of the null hypotheses are possible by performing additional contingency analyses nested within the original contingency table (Templeton 1987, 1996). The first nested series examined the distribution of nonsynonymous and synonymous mutations in the transmembrane *vs.* surface domains. None of the contingency tests involving synonymous mutations resulted in a rejection of the null hypothesis, nor did the tests for nonsynonymous mutations. However, there are very few nonsynonymous mutations overall (especially in the surface regions), so this lack of significance could be due to the much lower statistical power in this case as compared to the synonymous mutation case (Templeton 1996). One way of regaining power in a contingency framework is to pool categories. The standard pooling of "polymorphic *vs.* fixed" (Templeton 1987, McDonald and Kreitman 1991) did not produce a significant result, nor did pooling into "young *vs.* old" (Templeton 1996).

The second nested series examined the distribution of nonsynonymous *vs.* synonymous mutations within the transmembrane and surface regions separately. The null hypothesis of neutrality is not rejected in the surface regions, but is strongly rejected in the transmembrane regions of the human *ND2* gene. An examination of the data reveals a significantly greater number of nonsynonymous polymorphisms (or deficiency of synonymous polymorphisms) within human transmembrane regions than expected based on interspecific comparisons. Within human transmembrane regions, 66.7% of the polymorphisms are nonsynonymous mutations. In contrast, between humans and chimpanzees, only 15.2% of the fixed differences are nonsynonymous mutations. Furthermore, there is an excess of nonsynonymous mutations in the young category (63.6%) compared with the old category (18.6%). Nested contingency analyses were not performed for the *cyt b* data since the null hypothesis of neutrality was not rejected in initial contingency tests (see Section 4.4.3.2).

The results presented here are generally consistent with other studies which have used DNA sequence data to test the hypothesis that mtDNA variation is neutral. In Drosophila, non-neutral patterns have been documented for *ND5* (Rand *et al.* 1994) and *cyt b* (Ballard and Kreitman 1994). Similar patterns have also been observed for *ND3* in mice (Nachman *et al.* 1994), and *ND3* (Nachman *et al.* 1996) and *COII* (Templeton

1996) in humans and chimpanzees. In all of these studies, the ratio of nonsynonymous to synonymous nucleotide differences is greater within species than between species (Table 4.10). What might account for the observed pattern?

TABLE 4.10 Summary of variation at nonsynonymous (NS) and synonymous (S) sites within and between different species.

		Fixed (Between)			Polymorphic (Within)		
Species	**Locus**	**NS**	**S**	**NS/S**	**NS**	**S**	**NS/S**
Drosophila	*ND5*[a]	15	54	0.28	7	11	0.64
	cyt b[b]	10	97	0.10	6	12	0.50
Mouse	*ND3*[c]	2	23	0.09	11	13	0.85
Chimpanzee	*ND3*[d]	4	31	0.13	4	3	1.33
	ND2[e]	11	84	0.13	7	32	0.22
	cyt b[e]	30 (25)	96 (110)	0.31 (0.23)	6	39	0.15
Human	*ND3*[d]	4	31	0.13	4	7	0.57
	ND2[e]	11	84	0.13	10	11	0.91
	cyt b[e]	30 (25)	96 (110)	0.31 (0.23)	7	19	0.37

NOTE.—With the exception of *cyt b* in chimpanzees, the ratio of nonsynonymous to synonymous polymorphisms within species is greater than the ratio for fixed differences between species. The tests are not significant for *ND5* in Drosophila, *ND2* and *cyt b* in chimpanzees, and *ND3* and *cyt b* in humans.

References: [a] Rand *et al.* (1994); [b] Ballard and Kreitman (1994); [c] Nachman *et al.* (1994); [d] Nachman *et al.* (1996); and [e] This study: the chimpanzee data is from the west African subspecies *P. t. verus*. For the *cyt b* data, both the minimum and maximum interspecific counts are shown (see Section 4.4.2.2).

One possible explanation for these observations is that some from of balancing selection is maintaining amino acid variability. This hypothesis is considered unlikely for the human data presented here because Tajima's (1989*a*) D_T test, and the D_F and F tests proposed by Fu and Li (1993) are significantly negative for all subsets of the *ND2* data. The majority of these tests are also significantly negative for the *cyt b* data, indicating an excess of low frequency polymorphisms (see Table 4.7). Under a model of balancing selection some polymorphisms would be maintained in the population at intermediate frequencies, thus leading to positive test values.

A second possible explanation for the results is that there has been a recent relaxation of selective constraint in the human lineage. This would result in some previously deleterious mutations becoming neutral and being incorporated into the population as polymorphism (Takahata 1993*b*). Takahata (1993*b*) has argued that deleterious mutations in the human population may have become harmless under the changed (improved) environment throughout the Pleistocene. However, this hypothesis again does not adequately explain the results of Tajima's and Fu and Li's tests. A relaxation of selective constraint is expected to increase the rate at which mutations at nonsynonymous sites are introduced into the population, but is expected to have very little such effect on mutations at synonymous sites. Under this scenario we would expect to observe negative test values for nonsynonymous sites but not for synonymous sites. However, the results presented here show that D_T, D_F, and F are negative for both nonsynonymous and synonymous sites. All of theses tests are significant for the human *ND2* data, and the majority are significant for the *cyt b* data (see Table 4.7). Negative test values have also been observed for the noncoding control region (Jorde *et al.* 1995).

The pattern of mtDNA variation is consistent with a model of a recent population bottleneck followed by an expansion in population size (*e.g.*, Di Rienzo and Wilson 1991; Rogers and Harpending 1992; Harpending *et al.* 1993; Sherry *et al.* 1994; Rogers and Jorde 1995). This model can be used to explain the negative values of Tajima's and Fu and Li's tests, however, it does not explain the contingency test results. Furthermore, if human mitochondrial genome diversity reflects historical patterns of population size change, then similar patterns are expected of nuclear genome diversity (see Section 1.5).

This appears not to be the case and differences in the patterns of mitochondrial and nuclear genome diversity have recently been interpreted as evidence against the population expansion scenario (Hey 1997). A population bottleneck in the human lineage also appears to be incompatible with the unusual polymorphism at the major histocompatibility complex loci (*e.g.*, Takahata 1990, 1993*a*; Klein *et al.* 1993; Ayala *et al.* 1994; Ayala 1995; Ayala and Escalante 1996), despite some criticism of the details of some of these analyses (Erlich *et al.* 1996). It is also inconsistent with the pattern of nucleotide polymorphism at the β-globin locus (Harding *et al.* 1997), and of *Alu* repeat and microsatellite variation (H. Harpending, personal communication).

Another possibility is that amino acid mutations at *ND2* are slightly deleterious (*e.g.*, Ohta 1992). Slightly deleterious mutants may persist within populations for brief periods, but they are unlikely to rise in frequency or become fixed. Slightly deleterious models of molecular evolution have previously been invoked as potential explanations for patterns of mitochondrial genome evolution in Drosophila (DeSalle and Templeton 1988; Ballard and Kreitman 1994), mice (Nachman *et al.* 1994) and humans (Nachman *et al.* 1996; Templeton 1996). The test results presented here reveal a significant excess of young nonsynonymous polymorphisms within human *ND2* transmembrane regions (see Section 4.4.3.1), suggesting that they may be deleterious. Selection against deleterious alleles maintained by mutation ("background selection") results in reduced levels of polymorphism at linked neutral sites (Charlesworth *et al.* 1993, 1995; Charlesworth 1994; Hudson 1994; Hudson and Kaplan 1994). For the case where all sites are completely linked, such as in the mitochondrial genome, the effect of background selection is to mimic a simple reduction in effective population size (Hudson 1994). The negative values of Tajima's and Fu and Li's tests are (in principle) consistent with this model. However, simulation studies and power analyses suggest that significant test values are rarely detected under a model of background selection (Hudson and Kaplan 1994; Charlesworth *et al.* 1995).

The relative contribution of slightly deleterious mutations to heterozygosity increases as effective population size, N_e, decreases (Kimura 1983). This phenomenon can be explained by the fact that when N_e is small, selection is not effective and all alleles behave almost as neutral alleles, so that even unfavourable alleles contribute significantly to heterozygosity (Li 1978). Thus, if nonsynonymous mutations in the mitochondrial genome are slightly deleterious we would expect a relative increase, compared to neutral synonymous mutations, as N_e decreases. The nonsynonymous to synonymous ratio in the *ND2* gene is significantly greater within humans than within chimpanzees (0.91:0.22, Table 4.10; FET probability = 0.0196). The results for the much smaller *ND3* gene are not significant (0.57:1.33, Table 4.10; FET probability = 0.6305) based on a comparison of 61 humans and five chimpanzees (Nachman *et al.* 1996).

Under neutrality, diversity should be low in small populations and high in large ones. Thus, the lower level of mitochondrial diversity in humans compared with chimpanzees (see Section 3.4.1 and Section 4.4.1; Ferris *et al.* 1981; Morin *et al.* 1994; Ruvolo *et al.* 1994; Nachman *et al.* 1996; Wise *et al.* 1997) may reflect a smaller N_e in humans. This is consistent with the slightly deleterious model presented above. However, the lower level of nuclear genome diversity in chimpanzees (see Section 3.4.2; Wise *et al.* 1997, and references therein) implies that the N_e of chimpanzees is smaller than humans, thus we would expect to observe a higher nonsynonymous/synonymous ratio within chimpanzees. This is inconsistent with the results presented here (see Table 4.10). The contingency test results could, however, reflect the occurrence of slightly deleterious mutations, if effective population size had been reduced for the mitochondrial genome by a selective sweep that did not affect most or all of the nuclear genome.

One frequent argument made in favour of the out-of-Africa replacement hypothesis is that the low levels of human mtDNA diversity imply, under neutrality, a small N_e during recent human evolutionary history (Rogers and Jorde 1995). However, directional selection could also explain the reduced mtDNA diversity in humans compared with chimpanzees. Because there is no apparent genetic recombination in mtDNA, this depletion of variation could be due to the hitchhiking effect of a selectively advantageous mutation sweeping to fixation in the recent evolutionary history of humans. As with

background selection, this would reduce the apparent size of N_e (Templeton 1993, 1994). The results of Tajima's and Fu and Li's tests (see Section 4.4.4) are consistent with the occurrence of directional selection in the human mitochondrial genome. In contrast to the slightly deleterious model, significant test values frequently occur in simulation studies for the hitchhiking model (Braverman *et al.* 1995; Simonsen *et al.* 1995). The disparate pattern of variation in the nuclear genome (Harding *et al.* 1997; Hey 1997) is also consistent with this explanation since, although the entire mitochondrial genome would be affected, the nuclear genome would be unaffected except possibly at specific genes interacting epistatically with mitochondrial genes.

The results presented here clearly demonstrate that the patterns of nucleotide sequence variation within the human mitochondrial *ND2* gene depart from neutral expectations. Patterns of variation within the *cyt b* gene are generally consistent with neutral expectations, but show similar trends to the *ND2* data. They confirm other studies that demonstrate apparent non-neutral evolution of the mitochondrial genome. The non-neutrality of mtDNA has severe implications for interpreting population parameters estimated from polymorphism data, such as effective population size. The evidence presented here suggests that evolutionary forces, such as background and/or directional selection, may lead to an incorrect interpretation of human mtDNA diversity levels. The low level of diversity may indicate the presence of selection, not small population size as previously thought.

Chapter 5

CONCLUSIONS AND FUTURE DIRECTIONS

Paleontology and archaeology are disciplines that traditionally deal with the reconstruction of human origins and history. They address questions such as when and where our species first emerged, how our ancestors spread over the globe and what major events occurred in their history (von Haeseler *et al.* 1995). During the past ten years, however, molecular genetics has come to make increasing contributions to this area. In particular, a famous paper by Cann *et al.* (1987*a*) examined mtDNA sequences from a global sample of humans. They suggested that the common (female) ancestor of these mitochondrial sequences originated in Africa some 200,000 years ago, and later dispersed from Africa, replacing other archaic human forms. Their conclusions seemed to provide strong support to one side of an acrid debate in human paleontology, and have stimulated much interest, research and debate concerning the origins of modern humans. The controversy resides not in the existence of a common mtDNA ancestor, but rather in the details concerning that ancestor and its implications (Templeton 1993).

5.1 THE AFRICAN ORIGIN

The African origin hypothesis of human mtDNA (Cann *et al.* 1987*a*; Vigilant *et al.* 1991) has been criticised on the basis of the tree reconstruction algorithms and the rooting methods (Maddison 1991; Hedges *et al.* 1992; Maddison *et al.* 1992; Templeton 1992, 1993). However, new tree-building methods (Penny *et al.* 1995; D'Andrade and Morin 1996; Strimmer and von Haeseler 1996) and more appropriate outgroups (Zischler *et al.* 1995; Krings *et al.* 1997) still place the root in Africa.

A variety of other genetic data also support an African origin of the human gene pool (see Section 1.4). For example, Nei and Takezaki (1996) analysed five different datasets based on protein polymorphisms, restriction fragment length polymorphisms, microsatellite allele frequencies, and *Alu* insertion elements in different human populations using chimpanzees as an outgroup. In all five cases, phylogenetic trees indicate an initial separation between African and non-African populations.

More extensive studies will eventually tell us if all, most, or just a fraction of our gene pool traces its ancestry back to Africa. The limited data available to date indicate that an African origin is likely for most genetic regions (but see Altheide and Hammer 1997; Hammer *et al.* 1997; Harding *et al.* 1997). However, as pointed out in Chapter 1, an African origin is compatible with both the multiregional and single origin (replacement) models of human evolution.

5.2 THE TIME OF ORIGIN

Analyses of mtDNA yield coalesence dates around 100,000 to 200,000 years ago (see Section 1.3). If these analyses are correct, then the implication is that non-African populations older than 200,000 years could not have contributed to the gene pool of modern humans. The archaic non-African populations would have been replaced, without much (if any) interbreeding, by anatomically modern populations spreading from Africa in the past few hundred thousand years (Stoneking *et al.* 1992). One limitation, of course, is that mtDNA is a single linked entity, the history of which does not necessarily reflect the history of the human species. Hypotheses built from mtDNA analyses must be tested with data from the nuclear genome. Nuclear data from protein polymorphisms (Nei and Roychoudhury 1974*a*), microsatellites (Goldstein *et al.* 1995*a*), *CD4* microsatellite haplotypes (Tishkoff *et al.* 1996) and Y chromosomes (Dorit *et al.* 1995; Hammer 1995) appear to agree with the dates established by the mtDNA data. Details of these studies, however, have been criticised for a variety of reasons (see Section 1.4). Furthermore, a recent study of β-globin sequence polymorphism suggests that modest differences in levels of diversity between Africa and Asia are better explained by greater African effective population size than greater time depth (Harding *et al.* 1997, see also Stoneking *et al.* 1997)). They conclude that there is no evidence for an exponential population expansion out of a bottlenecked founding population 200,000 years ago.

5.3 COMPARISON OF MITOCHONDRIAL AND NUCLEAR DATA

Other studies show that mitochondrial and nuclear data present discordant pictures of human origins (see Section 1.5; Jorde *et al.* 1995; Hey 1997). To further address the

generality of this result we examined the relative levels of mitochondrial and nuclear genome diversity in humans and other closely related species, particularly chimpanzees (see Chapter 3; Wise *et al.* 1997).

When mitochondrial control region sequences are compared in 1,554 geographically diverse humans, the mean pairwise sequence difference (π) is 2.0%. In contrast, 115 chimpanzees differ by an average of 7.5% in the same region. Similarly, studies of restriction site polymorphisms and DNA sequences in the mitochondrial genomes of humans, chimpanzees, gorillas and orangutans have shown humans to be much less diverse than apes (Ferris *et al.* 1981; Ruvolo *et al.* 1994). This does not seem to be due to a lower mutation rate in humans (Blanchard and Schmidt 1996). Thus, the data suggest that human mtDNA variation is about three times younger than that of chimpanzees. However, great ape species differ from humans in that they appear to consist of reproductively isolated subpopulations corresponding to recognised subspecies. This is illustrated in Figure 3.1 which shows that chimpanzee subspecies correspond to distinct clades on the mtDNA phylogenetic tree.

It has been suggested that intersubspecific genetic differences may account for the observed large genetic diversity within ape species (Goldberg and Ruvolo 1997). Indeed, genetic distances between subspecies are known to be large in chimpanzees, gorillas and orangutans (Morin *et al.* 1994; Ruvolo *et al.* 1994; Garner and Ryder 1996). Examination of the data presented here for the mitochondrial control region, *ND2* gene and *cyt b* gene (see Sections 3.4.1 and 4.4.1) suggests that nucleotide diversity within western chimpanzees is two to three times higher than within the entire human species. Control region diversity for a smaller sample of central chimpanzees is two-fold higher (see Section 3.4.1), whereas a larger eastern chimpanzee sample shows similar levels to humans (Goldberg and Ruvolo 1997). Furthermore, control region diversity for a small sample of western lowland gorillas is also very high (π = 7.2%), based on sequences reported in Garner and Ryder (1996). It therefore appears unlikely that the relatively high levels of mitochondrial genome diversity within ape species is entirely due to genetic divergence between subspecies.

A recent common origin is widely accepted as the cause of low mitochondrial genome diversity in humans (Cann *et al.* 1987*a*; Vigilant *et al.* 1991). A Late Pleistocene population bottleneck has specifically been implicated as the key event marking the genetic origin of modern human populations (Harpending *et al.* 1993; Rogers 1995). Changes in population size and structure should be similarly reflected in the patterns of nuclear genome diversity. Thus, if neutrality is assumed, humans should also have low levels of nuclear genome diversity relative to other primate species. However, based on heterozygosity (H) at protein-coding and tandem repeat loci, humans appear to have higher levels of nuclear genome diversity than chimpanzees (see Section 3.4.2; Wise *et al.* 1997 and references therein). Furthermore, the ratio of mitochondrial to nuclear genome diversity (π/H) in humans is substantially less than in other catarrhines (see Table 3.3), implying that humans, not chimpanzees, have an unusual π/H ratio.

This discrepancy could result, under a neutral model, from either different mutation rates or differences in the population structure and history (*e.g.*, sex ratio, population subdivision, migration) (see Section 3.5). For example, low levels of migration among human populations could affect nuclear and mtDNA variation differently. At the appropriate level, limited migration could allow nuclear variation to be maintained, but given the lower effective population size of mtDNA (due to the haploid and maternal mode of inheritance), its effective migration rate could be sufficiently low to allow drift within local populations to reduce mtDNA variation (Hale and Singh 1991). However, it is unlikely that such a scenario is the sole explanation for the discrepancy. A thorough theoretical investigation is required to determine the possible ways in which these factors might have interacted to produce the low human π/H ratio.

Natural selection provides an alternative explanation for the apparently inconsistent pattern of mtDNA variation in humans. There is increasing evidence to suggest that the mitochondrial genome does not evolve according to a strictly neutral model of evolution (*e.g.*, Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Templeton 1996; Hey 1997; Wise *et al.* 1998). A variety of selective forces can mimic the effects of interpopulation migration or population bottlenecks.

5.4 NON-NEUTRAL EVOLUTION OF THE MITOCHONDRIAL GENOME

To further investigate selection on the mitochondrial genome we examined sequence variation in the *ND2* and *cyt b* genes of humans and chimpanzees (see Chapter 4; Wise *et al.* 1998). Patterns of variation in humans depart from neutrality by several test criteria. First, a comparison within and between species of the ratio of nonsynonymous to synonymous nucleotide differences reveals a relative excess of nonsynonymous polymorphisms within humans compared to the neutral prediction (McDonald and Kreitman 1991). This pattern of excess nonsynonymous polymorphism is not seen within chimpanzees. Second, there is evidence for an excess of unique or low frequency polymorphisms (Tajima 1989*a*), as well as an excess of young mutations (Fu and Li 1993) in the human data, but not in the chimpanzee data.

Three hypotheses are compatible with the McDonald and Kreitman (1991) test results: balancing selection, relaxation of selective constraints and slightly deleterious mutations. The Tajima (1989*a*) and Fu and Li (1993) test results are consistent with either a population bottleneck or slightly deleterious mutations (see Section 4.5). By combining the results from these two analyses, one possible explanation for the reduced mtDNA variation in humans is background selection against newly arising deleterious mutations. However, it is difficult to distinguish between background selection and genetic hitchhiking associated with directional selection (a selective sweep) as explanations for reduced variation in tightly linked regions. It has been suggested that significant Tajima's and Fu and Li's tests are rarely detected under a model of background selection unless the selection against deleterious mutations is extremely weak (Hudson and Kaplan 1994; Charlesworth *et al.* 1995). On the otherhand, these tests have reasonable power to detect the excess of rare polymorphisms predicted by selective sweeps. (Braverman *et al.* 1995; Simonsen *et al.* 1995). Thus, the results presented here (see Section 4.4.4) suggest that the lack of mtDNA variation in humans may best be explained by a recent selective sweep, followed by the accumulation of slightly deleterious mutations that result in the observed relative higher ratio of nonsynonymous to synonymous mutations. As illustrated in Figure 5.1, the presence of selection makes it difficult to distinguish between alternative models of human evolution when considering a single genetic locus.

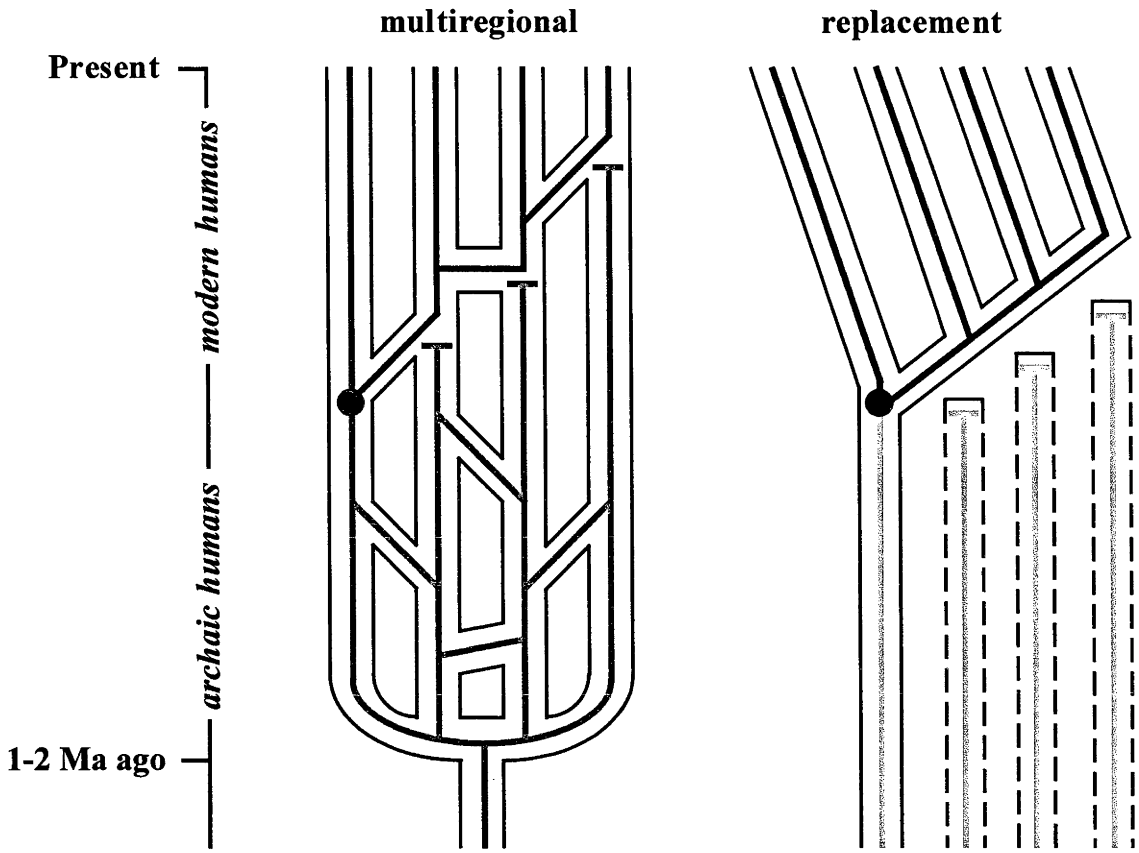


FIGURE 5.1 A possible scenario of human mtDNA evolution involving an advantageous mtDNA type sweeping through the population (blue), replacing existing types (pink). Under this scenario the mtDNA data are consistent with multiregional evolution. This pattern would be similar to one produced if modern humans evolved from a single isolated population of archaic humans (green) and spread across the globe replacing all other archaic human populations (yellow).

5.5 MITOCHONDRIA AND HUMAN EVOLUTION

Variation in the mitochondrial genome has been the major source of evidence for the recent spread of humans from Africa and their replacement of pre-existing human forms throughout the Old World. These studies assume that the mtDNA data reflects the movements and fates of entire human populations and not just the movements and fates of specific mtDNA types. Hitchhiking is the spread of a selectively favoured mutation through a population on variation at closely linked sites (Maynard-Smith and Haigh

1974). The process where the site which is under positive selection pressure and all linked sites becomes fixed in a population is termed selective sweep (Berry *et al.* 1991).

The most significant change in our recent evolutionary past has been an enormous increase in the complexity of our brain and its subsequent energy requirements. Energy in the body is generated by the mitochondria, and many genes involved in the process are encoded in the mitochondrial genome (see Section 1.2). These genes are therefore potential candidates for the evolution of complex brain function. This is supported by the fact that many genetic diseases involving mitochondrial genes result in neurological disorders (*e.g.*, Wallace 1992, 1995).

As discussed above, the pattern of variation in the human mitochondrial genome is not consistent with what is predicted by neutral evolutionary models, nor is it consistent with the patterns in the nuclear genome or patterns seen in chimpanzees. Since mitochondria appear to play a critical role in central nervous system function it is feasible to expect a favourable mutation could convey a selective advantage to the individual and therefore any neutral variants associated with it. For example, as human brain size began to increase rapidly during the Middle Pleistocene, there may have been selective pressure for superior mitochondrial genes capable of satisfying the increased energy demands of a larger brain. This is currently only speculation and more evidence is needed before it can be firmly established.

5.6 FUTURE DIRECTIONS

The available genetic data suggests that the mitochondrial and nuclear genomes of humans have different evolutionary histories (Hey 1997, Wise *et al.* 1997). Complex models need to be developed to investigate how various population parameters might have interacted to produce this discrepancy. One simple explanation is that mitochondrial genome evolution has been shaped by natural selection (Nachman *et al.* 1996; Templeton 1996; Hey 1997; Wise *et al.* 1998). If this is true, then the history of modern human populations cannot be accurately inferred from mitochondrial data. Models of human evolution must be tested with data from the nuclear genome. To date, most studies have

focused on immune system loci, microsatellites, *Alu* insertions and the Y-chromosome, all of which have potential problems (see Chapter 1). It is therefore important to identify appropriately variable segments throughout the nuclear genome. One method is to look at sequence variation in a number of different loci. The X-chromosome provides an attractive starting point because males are hemizygous, avoiding the problems associated with sequencing diploid DNA (Hey 1997). The Human Genome Project, with its goal of determining the complete sequence of the human genome, should greatly facilitate such studies. Comparative studies of nuclear (sequence) variation in humans and non-human primates are also required to further characterise the human genome and trace its evolution.

APPENDICES

APPENDIX A

Nucleotide sequence alignment of the first hypervariable segment of the mitochondrial control region from 115 common chimpanzees. Nucleotide positions are numbered according to the human sequence "ANDERSON" (Anderson *et al.* 1981). Additional nucleotides not found in this sequence are ignored in the numbering system. Dots (.) indicate sequence identity, dashes (-) denote insertions/deletions, and ? denotes missing data. Sequences reported in this study have the prefix "A"; sequences obtained from Kocher and Wilson (1991) are C1 to C3; and sequences obtained from Morin *et al.* (1994) are named according to that study.

A-176 through to WL3 are from the east African subspecies *Pan troglodytes schweinfurthii*, A-60 through to Ptt62 are from the central African subspecies *P. t. troglodytes*, and A-26 through to Tobar are from the west African subspecies *P. t. verus* (see Figure 3.1).

```
           16024       16034       16044       16054       16064       16074
           |           |           |           |           |           |
A-176      TTCTTTCATG GGGAAGCAAA TTTAAGTACC ACCTAAGTAT TGGCTTATTC ATTA-CAACC
Al2        ?????????? ?????????? ???????... .......??? ??........ ..........
Be1        ?????????? ?????????? ?????????? ?????????? ????????.. ..........
California ?????????? ?????????? ?????????? ?????????? ?????????. ..........
DH1        ?????????? ?????????? ?????????? ?????????? ????????.. ..........
Evered     ?????????? ?????????? ??????.... .......... .......... ..........
Frodo      ?????????? ?????????? ??????.... .......... .......... G.........
George     ?????????? ?????????? ?????????? .......??? ??........ ..........
Georgette  ?????????? ?????????? ?????????? ?????????? ?????????. .........?.
Gigi       ?????????? ?????????? ?????????? ?????????? ??........ ..........
Gimble     ?????????? ?????????? ????...... .......... .......... ..........
JF1        ?????????? ?????????? ?????????? ??.....??? ?......C.. ........?..
Kidevu     ?????????? ?????????? ?????????? .......??? ???..... G.........
Maggie     ?????????? ?????????? ?????????? ?????????? ??........ ..........
Max        ?????????? ?????????? ?????????? ?????????? ?????????. ..........
Me1        ?????????? ?????????? ?????????? ?????????? ????????.. ..........
Noel       ?????????? ?????????? ?????????? ?????????? ?????????? G........?.
Pal        ?????????? ?????????? ?????????? ?????????? ??........ ..........
Prof       ?????????? ?????????? ????...... .......... .......C.. ..........
Pts2       ?????????? ?????????? ?????????? ?????????? ??????.... ..........
Pts6       ?????????? ?????????? ?????????? ?????????? ?????????? ????.?????
Pts12      ?????????? ?????????? ?????????? ?????????? ?????????? ..........
Pts20      ?????????? ?????????? ?????????? ?????????? ?????????? ??........
Pts21      ?????????? ?????????? ?????????? ?????????? ?????????? ..........
Pts29      ?????????? ?????????? ?????????? ?????????? ?????..... ..........
Pts35      ?????????? ?????????? ?????????? ?????????? ?????????? .....?????
Pts43      ?????????? ?????????? ?????..... ........?? .......... ..........
Pts49      ?????????? ?????????? ?????????? ?????????? ?????????? ..........
Pts50      ?????????? ?????????? ?????????? ?????????? ??........ ..........
Pts51      ?????????? ?????????? ?????????? ?????????? ?????..... ..........
Ruhara     ?????????? ?????????? ?????????? ?????????? ???????... ..........
Sandi      ?????????? ?????????? ?????????? ?????????? ?????????? ??........
Skosha     ?????????? ?????????? ?????????? ?????????? ??........ ..........
Spindle    ?????????? ?????????? ??????.... .......... .......... ..........
Sy2        ?????????? ?????????? ?????????? ????...??? ??????.... ..........
TT3        ?????????? ?????????? ?????????? ?????????? ?????????. ..........
Tubi       ?????????? ?????????? ???????... .......??. ?...????.. ..........
WL3        ?????????? ?????????? ????????.. .......??? ???..... G.........

A-60       .......... .......... .......... .......... .......... ..........
A-175      .......... .......... .......... .......... .......... ..........
C1         .......... .......... .......... .......... .......... ..........
Berthe     ?????????? ?????????? ?????..... .......... ....C..... ..........
Gemini     ?????????? ?????????? ?????..... .......... .......... ..........
Ivindo     ?????????? ?????????? ?????..... .......... .......... ..........
Makokou    ?????????? ?????????? ?????..... .......... .......... ..........
Milla      ?????????? ?????????? ?????.?... .......??? ??..C..... .......?..
Mopia      ?????????? ?????????? ?????..... .......... .......... ..........
Mouilla    ?????????? ?????????? ?????..... .......... .......... ..........
Ntebe      ?????????? ?????????? ?????..... .......... .......... ..........
Ponia      ?????????? ?????????? ?????..... .......... .......... ..........
Ptt33      ?????????? ?????????? ????...... .......... .?........ ..........
Ptt34      ?????????? ?????????? ?????????? ??.....??? ??........ ..........
Ptt35      ?????????? ?????????? ?????????? ?????????? ?????????? ????.?????
Ptt36      ?????????? ?????????? ?????..... .......... ?...C..... ..........
Ptt37      ?????????? ?????????? ????...... .......... ?......... ..........
Ptt43      ?????????? ?????????? ?????????? ??.....??? ??........ ..........
Ptt59      ?????????? ?????????? ?????????? ?????????? ??........ ..........
Ptt62      ?????????? ?????????? ?????????? ?????????? ?????????? ..........
```

```
            16024       16034       16044       16054       16064       16074
            |           |           |           |           |           |
A-26        TTCTTTCATG GGGAAGCAAA TTTAAGTACC ACCTAAGTAC TGGCTCATTC ATTA-CAACC
A-33        .......... .......... .......... .......... .......... ..........
A-42        .......... .......... .......... .......... .......... ..........
A-50        .......... .......... .......... .......... .......... ..........
A-52        .......... .......... .......... .......... .......... ..........
A-56        .......... .......... .......... .......... .......... ..........
A-62        .......... .......... .......... .......... .......... ..........
A-69        .......... .......... .......... .......... .......... ..........
A-89        .......... .......... .......... .......... .......... ..........
A-90        .......... .......... .......... .......... .......... ..........
A-94        .......... .......... .......... .......... .......... ..........
A-97        .......... .......... .......... .......... .......... ..........
A-101       .......... .......... .......... .......... .......... ..........
A-102       .......... .......... .......... .......... .......... ..........
A-108       .......... .......... .......... .......... .......... ..........
A-118       .......... .......... .......... .......... .......... ..........
A-128       .......... .......... .......... .......... .......... ..........
A-129       .......... .......... .......... .......... .......... ..........
A-130       .......... .......... .......... .......... .......... ..........
A-131       .......... .......... .......... .......... .......... ..........
A-136       .......... .......... .......... .......... .......... ..........
A-137       .......... .......... .......... .......... .......... ..........
A-139       .......... .......... .......... .......... .......... ..........
A-172       .......... .......... .......... .......... .......... ..........
A-179       .......... .......... .......... .......... .......... ..........
A-182       .......... .......... .......... .......... .......... ..........
A-192       .......... .......... .......... .......... .......... ..........
A-197       .......... .......... .......... .......... .......... ..........
A-199       .......... .......... .......... .......... .......... ..........
A-208       .......... .......... .......... .......... .......... ..........
A-230       .......... .......... .......... .......... .......... ..........
A-231       .......... .......... .......... .......... .......... ..........
A-235       .......... .......... .......... .......... .......... ..........
A-239       .......... .......... .......... .......... .......... ..........
A-242       .......... .......... .......... .......... .......... ..........
A-268       .......... .......... .......... .......... .......... ..........
A-281       .......... .......... .......... .......... .......... ..........
A-282       .......... .......... .......... .......... .......... ..........
A-283       .......... .......... .......... .......... .......... ..........
A-284       .......... .......... .......... .......... .......... ..........
A-285       .......... .......... .......... .......... .......... ..........
A-286       .......... .......... .......... .......... .......... ..........
A-288       .......... .......... .......... .......... .......... ..........
A-290       .......... .......... .......... .......... .......... ..........
A-291       .......... .......... .......... .......... .......... ..........
A-292       .......... .......... .......... .......... .......... ..........
C2          .......... .......... .......... .......... .......... ..........
C3          .......... .......... ....G..... .......... .......... ..........
Edgar       ?????????? ?????????? ?????????? ?????????? ?????????. ..........
NC588       ?????????? ?????????? ?????????? .??....??? ??.?.?.... ..........
P1          ?????????? ?????????? ?????????? ?????????? ?......... ..........
P2a         ?????????? ?????????? ?????????? ?????????? ?......... ..........
P3          ?????????? ?????????? ?????????? ?????????? ?......... ..........
Ptv4        ?????????? ?????????? ??????.... .......... .........C. ..........
Ptv5        ?????????? ?????????? ?????????? ?......... .......... ..........
Ptv15       ?????????? ?????????? ??????.... .......... .......... ..........
Tobar       ?????????? ?????????? ????.A.... .......... .......... .......??.
ANDERSON    .......... ........G. ...GG..... ...C.....T ..A....CC. ..C.A.....
```

```
            16084       16094       16104       16114       16124       16134
            |           |           |           |           |           |
A-176       ACTATGTATT TCGTACATTA CTGCTAGCCA CCATGAATAT TGTACAGTAC TATAATCACT
Al2         G.......C. .T........ ....C..... .......... .......... ..........
Be1         G.......C. .T........ ....C..... .......... .......... ..........
California  G......... .......... ....C..... .......... .......... ........T.
DH1         G...?..... .......... ....C..... .......... .......??. ....?.....
Evered      G......... .......... ....C..... .......... .......... ........T.
Frodo       G......... .......... ....C..... .......... .......... ..........
George      G......... .......... ....C..... .......... .......... ..........
Georgette   G......... .......... ....C..... .......... .......... ..........
Gigi        G......... .......... ....C..... .......... .......... ........T.
Gimble      G......... .......... ....C..... .......... .......... ..........
JF1         G......... .......... ....C..... .......... .......... ..........
Kidevu      G......... .......... ....C..... .......... .......... ..........
Maggie      G......... .......... ....C..... .......... .......?.. ..........
Max         G......... .......... ....C..... .......... .......... ..........
Me1         G.......C. .......... ....C..... .......... .......... ..........
Noel        G......... .......... ....C..... .......... .......... .........C
Pal         G......... .......... ....C..... .......... ........?. ..........
Prof        G......... .......... ....C..... .......... .......... ..........
Pts2        G......... .......... ....C..... .......... .......... ..........
Pts6        ?????..... .......... ....C..... .......... .......... ..........
Pts12       G......... .......... ....C..... .......... .......... ..........
Pts20       G......... .......... ....C..... .......... .......... ..........
Pts21       G......... .......... ....C..... .......... .......... ..........
Pts29       G......... .......... ....C..... .......... .......... ..........
Pts35       G.......C. .......... ....C..... .......... .......??. ..C.......
Pts43       G......... .......... ....C..... .......... .......... ..........
Pts49       G......... .......... ....C..... .......... .......... ..........
Pts50       G......... .......... ....C..... .......... ..?...???. ..........
Pts51       G......... .......... ....C..... .......... .......... ....?.....
Ruhara      G......... .......... ....C..... .......... .......... ..........
Sandi       G.......C. .......... ....C..... .......... .......... ..?.......
Skosha      G...?..... .......... ....C..... .......... .......... ........T.
Spindle     G......... .......... ....C..... .......... .......... ........T.
Sy2         G.......C. .......... ....C..... .......... .......... ..........
TT3         G......... .......... ....C..... .......... .......... .........T.
Tubi        G.....?... .......... ....C..... ......?... ......???? ?????...T.
WL3         G......... .......... ....C..... .......... .......?.. ..........

A-60        G......... .......... ....C..... .......... .......... ..........
A-175       G......... .......... ....C..... .......... .......... ..........
C1          G......... .......... ....C..... .......... .......... ..........
Berthe      G......... .......... ....C..... .......... .......... ..........
Gemini      G......... .......... ....C..... .......... .......... ..........
Ivindo      G......... .......... ....C..... .......... .......... ..........
Makokou     G........C .......... ....C..... .......... .......... ..........
Milla       G......... .......... ....C..... .......... .......... .........C
Mopia       G......... .......... ....C..... .......... .......... ..........
Mouilla     G......... .......... ....C..... .......... .......... ..........
Ntebe       G......... .......... ....C..... .......... .......... .........C
Ponia       G......... .......... ....C..... .......... .......... ..........
Ptt33       G......... .......... ....C..T.. .......... .......... ..........
Ptt34       G......... .......... ....C..T.. .......... .......... ..........
Ptt35       ???....... .?.?.?.... ?..??..??. ??........ .......... ..........
Ptt36       G......... .......... ....C..... .......... .......... ..........
Ptt37       G......... .......... ....C..... .......... .......... .........C
Ptt43       G......... .......... ....C..... .......... .......... .........C
Ptt59       G......... .......... ....C..... .......... .......... ..........
Ptt62       G......... .......... ....C..... .......... .......... ..........
```

```
          16084      16094      16104      16114      16124      16134
          |          |          |          |          |          |
A-26      GCTATGTATT TCGTACATTA CTGCCAGCCA CCATGAATAT TGTACAGTAC CATAATCACC
A-33      .......... .......... .......... .......... .......... ..........
A-42      .......... .......... .......... .......... .....G.... ......T...
A-50      .......... .......... .......T.. .......... .......... ..........
A-52      .......... .......... .......T.. .......... .......... ..........
A-56      .......... .......... .......... .......... .......... ..........
A-62      .......... .......... .......... .......... .......... ..........
A-69      .......... .......... .......... .......... .......... ..........
A-89      .......... .......... .......... .......... .....G.... ......T...
A-90      .......... .......... .......... .......... .......... ..........
A-94      .......... .......... .......... .......... C......... ..........
A-97      .......... .......... .......... .......... .......... .........T
A-101     .......... .......... .......... .......... .....G.... ..........
A-102     .......... .......... .......... .......... C......... ..........
A-108     .......... .......... .......... .......... .......... ..........
A-118     .......... .......... .......... .......... .....G.... ......T...
A-128     .......... .......... .......... .......... C......... ..........
A-129     .......... .......... .......... .......... .....G.... ..........
A-130     .......... .......... .......... .......... C......... ..........
A-131     .......... .......... .......... .......... .....G.... ......T...
A-136     .......... .......... .......... .......... .....G.... ......T...
A-137     .......... .......... .......... .......... .......... ..........
A-139     .......... .......... .......... .......... .......... ..........
A-172     .......... .......... .......... .......... .......... ..........
A-179     .......... .......... .......... .......... .......... ..........
A-182     .......... .......... .......... .......... .....G.... ..........
A-192     .......... .......... .......... .......... .......... ..........
A-197     .......... .......... .......... .......... .......... ..........
A-199     .......... .......... .......T.. .......... .......... ..........
A-208     .......... .......... .......... .......... C......... ..........
A-230     .......... .......... .......... .......... .....G.... ......T...
A-231     .......... .......... .......... .......... C......... ..........
A-235     .......... .......... .......... .......... .....G.... ..........
A-239     .......... .......... .......T.. .......... .......... ..........
A-242     .......... .......... .......... .......... .......... ..........
A-268     .......... .......... .......... .......... .......... ..........
A-281     .......... .......... .......... .......... .......... ..........
A-282     .......... .......... .......... .......... .......... ..........
A-283     .......... .......... .......... .......... .......... ..........
A-284     .......... .......... .......... .......... .......... ..........
A-285     .......... .......... .......... .......... C......... ..........
A-286     .......... .......... .......... .......... .......... ..........
A-288     .......... .......... .......... .......... .......... ..........
A-290     .......... .......... .......... .......... .......... ..........
A-291     .......... .......... .......... .......... .....G.... ......T...
A-292     .......... .......... .......... .......... .......... ..........
C2        .......... .......... .......... .......... .....G.... ......T...
C3        .......... .......... .......... .......... C......... .....-.....
Edgar     .......... .......... .......... .......... .????????? T...?.....
NC588     .......... .......... .......... .......... .......... ..........
P1        .......... .......... .......... .......... .......... ..........
P2a       .......... .......... .......... .......... .....G.... ..........
P3        .......... .......... .......... .......... .....G.... ......T...
Ptv4      .......... .......... .......... .......... .....G.... ......T...
Ptv5      .......... .......... .......... .......... .......... ..........
Ptv15     .......... .......... .......... .......... .......... ..........
Tobar     .......... .......... .......... .....?.... .....G.??G .?????????
ANDERSON  .......... .......... .......... .......... .....G.... .....AT..T
```

```
                16144       16154       16164       16174       16182       16192
                |           |           |           |           |           |
A-176           CAACCACCTA  TAATACATTA  AA-CCCACCC  C--ACATTAC  AACCTTCACC  CCATGCTTAC
Al2             ....T.....  ........C.  ....T.....  ..........  .....C....  ..........
Be1             ....T.....  ........C.  ....T.....  ..........  .....C....  ..........
California      ....T.....  ........C.  ..........  ..........  .....C....  ..........
DH1             ....T.....  ........C.  ..........  .........T  .....CT...  .........?.
Evered          ....T.....  ........C.  ..........  ..........  .....C....  ..........
Frodo           ....T.....  ........C.  ..........  .........T  .....C....  ..........
George          ....T.....  ........C.  ..........  ..........  .....C....  ..........
Georgette       ....T.....  ........C.  ..........  .......C.T  .....CT...  ..........
Gigi            ....T.....  ........C.  ..........  .......C..  .....C....  ..........
Gimble          ....T.....  ........C.  ..........  .........T  .....C....  .T........
JF1             ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Kidevu          ....T.....  ........C.  ..........  .........T  .....C....  ..........
Maggie          ....T.....  ........C.  ...T......  ..........  .....C....  ..........
Max             ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Me1             ....T?....  ........C.  ....T.....  ..........  .....C....  ..........
Noel            ....T.....  ........C.  ..........  .........T  .....C....  ..........
Pal             ....T.....  ........C.  ..........  .........T  .....C....  ..........
Prof            ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts2            ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts6            ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts12           ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts20           ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts21           ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts29           ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts35           ....T.....  ........C.  ..........  .........T  .....C....  ..........
Pts43           ....T.....  ........C.  ..........  .........T  ....??T...  ..........
Pts49           ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts50           ....T.....  ........C.  ..........  ..........  .G...C....  ..........
Pts51           ....T.....  ........C.  ..........  ..........  .G...C....  .........??
Ruhara          ....T.....  ........C.  ..........  ..........  .....C....  .T........
Sandi           ....T.....  ........C.  ....T.....  ..........  .....C....  ..........
Skosha          ....T.....  ........C.  ....T.....  .......C..  .....C....  ..........
Spindle         ....T.....  ........C.  ..........  .......C..  .....C....  ..........
Sy2             ....T.....  ........C.  ....T.....  ..........  .....C....  ..........
TT3             ....T.....  ........C.  ..........  ..........  .....C....  ..........
Tubi            ....T.....  ........C.  ..........  .......C..  .....C....  ..........
WL3             ....T.....  ........C.  ..........  .........T  .....C....  ..........

A-60            ....T.....  ...C......  ........T.  ..........  ...G.C.C..  ..........
A-175           ....T.....  ...C......  ........T.  ..........  ...G.C.C..  ..........
C1              ....T.....  ...C....C.  .......AT.  .C........  ...G.C.C..  ..........
Berthe          ....T.....  ..........  ..........  .C.....A..  .....C....  .T........
Gemini          ....T.....  ..GC......  ........T.  ..........  ...G.C.C..  ..........
Ivindo          ....T.....  ...C......  ...T....T.  ..........  ..TG.C.C..  ..CC......
Makokou         ....T.....  ...C......  ..........  ..........  ...G.C.C..  T.........
Milla           ....T.....  ..........  ..........  .TT.......  .....C....  .T........
Mopia           ....T.....  ...C......  ..........  ..........  ...G.C.C..  ..........
Mouilla         ....T.....  ...C......  ........T.  ..........  ...G.C.C..  ..........
Ntebe           ..........  ..........  ..........  ..........  ...G.C....  ..........
Ponia           ....T.....  ..........  ........T.  ..........  ...G.C.C..  ..........
Ptt33           ....T.....  ...C......  ........T.  ..........  ...G.C.C..  ..........
Ptt34           ....T.....  ...C......  ........T.  ..........  ...G.C.C..  ..........
Ptt35           ....T.....  ...C....C.  ........T.  ..........  ..TG.C.C..  ..........
Ptt36           ....T.....  ..........  ..........  .CC.......  .....C....  .T........
Ptt37           ....T.....  ...C......  ..........  ..........  ...G.C.C..  ..........
Ptt43           ..........  ..........  ..........  ..........  ...G.C....  ..........
Ptt59           ....T.....  ...C......  .....T..T.  ..........  ...G.C.C..  ..........
Ptt62           ....T.....  ...C......  .?........  ..........  ...G?C.C..  ..........
```

```
          16144      16154      16164      16174      16182      16192
          |          |          |          |          |          |
A-26      CAACTACCTA TAACACATAA AA-TCCACTC CC-ACACCAA AACCTTCACC CCATGCTTAC
A-33      ....C..... ..G....... ........CT .-....T... .......... ..........
A-42      .......... .......... ...C....C. .-....TT.. .......... ..........
A-50      .......... ...T...... .......... .......... .......... ..........
A-52      .......... ...T...... .......... .......... .......... ..........
A-56      ....C..... .........G ........CT .-....TT.. .......... ..........
A-62      ....C..... ...T...... ...C....C. .-....TT.. .......... ..........
A-69      .......... .......... .......... .......... .......... ..........
A-89      .......... .......... ...C....C. .-....TT.. .......... ..........
A-90      ....C..... ..G....... ........CT .-....TT.. .......... ..........
A-94      .......... ..GT...... .......... ......T... G......... ..........
A-97      ....C..... ..G....... ........CT .-....TT.. .......... ..........
A-101     .......... ..G....... ...C....C. .-....TT.. .......... ..........
A-102     .......... ..GT...... .......... ......T... G......... ..........
A-108     .......... ...T...... .......... .......... .......... ..........
A-118     .......... .......... ...C....C. .-....TT.. .......... ..........
A-128     .......... ..GT...... .......... ......T... G......... ..........
A-129     .......... ..G....... ...C....C. .-....TT.. .......... ..........
A-130     .......... ..GT...... .......... ......T... G......... ..........
A-131     .......... .......... ...C....C. .-....TT.. .......... ..........
A-136     .......... .......... ...C....C. .-....TT.. .......... ..........
A-137     ....C..... ..G......G .........T .-....TT.. .......... ..........
A-139     ....C..... ..G....... ........CT .-....TT.. .......... ..........
A-172     .......... .......... ...C....C. ......T... .....C.... T.........
A-179     .......... .......... .......... .......... .......... ..........
A-182     ....C..... ...T...... ...C....C. .-....TT.. .......... ..........
A-192     .......... .......... ...C....C. ......T... .....C.... T.........
A-197     .......... .......... ...C....C. ......T... .....C.... T.........
A-199     .......... ...T...... .......... .......... .......... ..........
A-208     .......... ..GT...... .......... ......T... G........T ..........
A-230     .......... .......... ...C....C. .-....TT.. .......... ..........
A-231     .......... ..GT...... .......... ......T... G........T ..........
A-235     .......... ..G....... ...C....C. .-....TT.. .......... ..........
A-239     .......... ...T...... .......... .......... .......... ..........
A-242     .......... .......... .......... .......... .......... ..........
A-268     .......... ...T...... .......... .......... .......... ..........
A-281     .......... .......... ...C....C. ......T... .....C.... T.........
A-282     .......... ..G....... ...C....C. .-....TT.. .......... ..........
A-283     .......... .......... .......... .......... .......... ..........
A-284     .......... .......... .......... .......... .....C.... T.........
A-285     .......... ..GT...... .......... ......T... G........T ..........
A-286     ....C..... ..G......G .........T .-....TT.. .......... ..........
A-288     ....C..... .........G ........CT .-....TT.. .......... ..........
A-290     .......... ...T...... .......... .......... .......... ..........
A-291     .......... .......... ...C....C. .-....TT.. .......... ..........
A-292     .......... .......... .......... .......... .......... ..........
C2        .......... .......... ...C...AC. ......TT.. .......... ..........
C3        .......... ..GT...... .......... ......T... .........T ..........
Edgar     .........? .......... .......... .......... .......... ..........
NC588     ....C..... ..G....... ........CT .-....TT.. .......... ..........
P1        .......... .......... .......... .......... .......... ..........
P2a       .......... ..G....... ...C....C. .-....TT.. .......... ..........
P3        .......... .......... ...C....C. .-....TT.. .......... ..........
Ptv4      .......... .......... ...C....C. .-....TT.. .G........ ..........
Ptv5      .......... .......... .......... .......... .......... ..........
Ptv15     .......... ...T...... .......... .......... .......... ..........
Tobar     ?????????? ?????????. ...C....C. .-....TT.. .......... ........??
ANDERSON  TG..C....G ..GT...... ..AC...A.. .-....T... ....CC.T.. ..........
```

```
            16202       16212       16222       16232       16242       16252
            |           |           |           |           |           |
A-176       AAGCACGCAC  AACAATCAAC  CCTCAACTGT  CATACATAAA  ACACAACTCC  AAAGACATTC
Al2         ..........  ..........  ..........  ..C.......  ..........  ..........
Be1         ..........  ..........  ..........  ..C.......  ..........  ..........
California  ..........  ..........  ..........  ..C.......  ..........  .......C..
DH1         ..........  ..........  ..........  ..C.......  ....T.....  ..........
Evered      ..........  ..........  ..........  ..C.......  ..........  .......C..
Frodo       ..........  ..........  ..........  ..C.......  ..........  ..........
George      ..........  ..........  ..........  ..C.......  ..........  ..........
Georgette   ..........  ..........  ..........  ..C.......  ..........  ..........
Gigi        ..........  ..........  ..........  ..C.......  ..........  .......C..
Gimble      ..........  ..........  ..........  ..C.......  ..........  ..........
JF1         ....?.....  ..........  ..........  ..C.......  ..........  .......C..
Kidevu      .........?  ??........  ..........  ..C.......  ..........  ..........
Maggie      ..........  ..........  ..........  ..C.......  ..........  ..........
Max         ..........  ..........  .T........  ..C.......  ..........  ..........
Me1         ..........  ..........  ..........  ..C.......  ..........  ..........
Noel        ..........  ..........  ..........  ..C.......  ..........  ..........
Pal         ..........  ..........  ..........  ..C.......  ..........  ..........
Prof        ..........  ..........  ..........  ..C.......  ..........  .......C..
Pts2        ..........  ..........  ..........  ..C.......  ..........  .......C..
Pts6        ..........  ..........  .T........  ..C.......  ..........  ..........
Pts12       ..........  ..........  .T........  ..C.......  ..........  ..........
Pts20       ..........  ..........  .T........  ..C.......  ..........  ..........
Pts21       ..........  ..........  .T........  ..C.......  ..........  ..........
Pts29       ..........  ..........  .T........  ..C.......  ..........  ..........
Pts35       ..........  ..........  ..........  ..C.......  ..........  ..........
Pts43       .....T....  ..........  ..........  ..C.......  ..........  .......C..
Pts49       ..........  ..........  ..........  ..C.......  ..........  .......C..
Pts50       ..........  ..........  .T........  ..C.......  ..........  ..........
Pts51       ..........  ..........  ..........  ..C.......  ..........  .......C..
Ruhara      ..........  ..........  ..........  ..C.......  ..........  ..........
Sandi       ..........  ..........  ..........  ..C.......  ..........  ..........
Skosha      ..........  ..........  ..........  ..C.......  ..........  .......C..
Spindle     ..........  ..........  ..........  ..C.......  ..........  .......C..
Sy2         ..........  ..........  ..........  ..C.......  ..........  ..........
TT3         ..........  ..........  ..........  ..C.......  ..........  .......C..
Tubi        ..........  ..........  ..........  ..C.......  ..........  .......C..
WL3         ..........  ..........  ..........  ..C.......  ..........  ..........

A-60        .......T..  ..........  ..........  ..C......G  ..G.......  .......C..
A-175       .......T..  ..........  ..........  ..C......G  ..G.......  .......C..
C1          .......T..  ..........  ..........  ..C.......  ..G.......  .......C..
Berthe      ..........  ..........  ..........  .GC.......  .TG.......  .......CCT
Gemini      .......T..  ..........  ..........  ..C.......  ..G.......  .......C..
Ivindo      .......T..  ..........  ..........  ..C......G  ..G.......  .......C..
Makokou     .......T..  ..........  ..........  ..C......G  ..G.......  .......C..
Milla       ....?.?...  ......????  ?.C.......  ..C.......  .TG.......  .......C..
Mopia       .......T..  ..........  ..........  ..C.......  ..G.......  .......C..
Mouilla     .......T..  ..........  ..........  ..C.......  ..G.......  .......C..
Ntebe       .....T.T..  ..........  ..........  ..C.......  ..G.......  ....T.....
Ponia       .....T.T..  ..........  ..........  ..C.......  ..G.......  ....G.....
Ptt33       .......T..  ..........  ..........  ..C.......  ..G.......  .......C..
Ptt34       .......T..  ..........  ..........  ..C.......  ..G.......  .......C..
Ptt35       .....T.T..  .G........  ..........  ..C.......  ..G.......  .......C..
Ptt36       ..........  ....G.....  ..C.......  ..C.......  .TG.......  .......CC.
Ptt37       ..........  ........G.  ..........  ..C.......  .TG.......  .......C..
Ptt43       .....T.T..  ..........  ..........  ..C.......  ..G.......  ....T.....
Ptt59       .......T..  ..........  ..........  ..C......G  ..G.......  .......C..
Ptt62       .......T..  ..........  ..........  ..C......G  ..G.......  .......C..
```

```
                16202       16212       16222       16232       16242       16252
                |           |           |           |           |           |
A-26            AAGCACGCAC  AACAATCAAC  CTCCAACTGT  CGAACATAAA  ACACAACTCC  AACGACACCC
A-33            ..........  ..........  .C......A.  ..........  ..........  ........TT
A-42            ..........  ..........  ...T....A.  .A........  ..........  ........TT
A-50            ..........  ..........  ..........  .A........  ..........  ..........
A-52            ..........  ..........  ..........  .A........  ..........  ..........
A-56            ..........  ..........  .C......A.  ..........  ..........  ........TT
A-62            ..........  ..........  ........A.  ..........  .......C..  ........T.
A-69            ..........  ..........  ..........  ..........  ..........  ..........
A-89            ..........  ..........  ...T....A.  .A........  ..........  ........TT
A-90            ..........  ..........  .C......A.  ..........  ..........  ........TT
A-94            ..........  ..........  TC........  ..........  ......T...  ..........
A-97            ..........  ..........  .C......A.  ..........  ..........  ........TT
A-101           .....T....  ..........  ...T....A.  .A........  ..........  .........T
A-102           ..........  ..........  TC........  ..........  ......T...  ..........
A-108           ..........  ..........  ..........  ..........  ..........  ..........
A-118           ..........  ..........  ...T....A.  .A........  ..........  ........TT
A-128           ..........  ..........  TC........  ..........  ......T...  ..........
A-129           ..........  ..........  ...T....A.  .A........  ..........  .........T
A-130           ..........  ..........  TC........  ..........  ......T...  ..........
A-131           ..........  ..........  ...T....A.  .A........  ..........  ........TT
A-136           ..........  ..........  ...T....A.  .A........  ..........  ........TT
A-137           ..........  ..........  .C......A.  ..........  ..........  ........TT
A-139           ..........  ..........  .C......A.  ..........  ..........  ........TT
A-172           ..........  ..........  .C........  ..........  ..........  ..T.......
A-179           ..........  ..........  ..........  ..........  ..........  ..........
A-182           ..........  ..........  ........A.  ..........  .......C..  ........T?
A-192           .......?..  ..........  .C........  ..........  ..........  ..T.......
A-197           .......?..  ..........  .C........  ..........  ..........  ..T.......
A-199           ..........  ..........  ..........  .A........  ..........  ..........
A-208           ..........  ..........  TC........  ..........  ......T...  ..........
A-230           ..........  ..........  ...T....A.  .A........  ..........  ........TT
A-231           ..........  ..........  TC........  ..........  ......T...  ..........
A-235           ..........  ..........  ...T....A.  .A........  ..........  .........T
A-239           ..........  ..........  ..........  .A........  ..........  ..........
A-242           ..........  ..........  ..........  ..........  ..........  ..........
A-268           ..........  ..........  ..........  ..........  ..........  ..........
A-281           ..........  ..........  .C........  ..........  ..........  ..T.......
A-282           ..........  ..........  ...T....A.  .A........  ..........  .........T
A-283           ..........  ..........  ..........  ..........  ..........  ..........
A-284           ..........  ..........  ..........  ..........  ..........  .........?
A-285           ..........  ..........  TC........  ..........  ......T...  ..........
A-286           ..........  ..........  .C......A.  ..........  ..........  ........TT
A-288           ..........  .........T  .C......A.  ..........  ..........  ........TT
A-290           ..........  ..........  ..........  ..........  ..........  ..........
A-291           ..........  ..........  ...T....A.  .A........  ..........  ........TT
A-292           ..........  ..........  ..........  ..........  ..........  ..........
C2              ..........  ..........  ...T....A.  .A........  ..........  ........TT
C3              ..........  ..........  TC........  ..........  ......T...  ..........
Edgar           ..........  ..........  ..........  ..........  ..........  ..........
NC588           ..........  .........?.  .C......A.  ..........  ..........  ........TT
P1              ..........  ..........  ..........  ..........  ..........  ......????
P2a             ..........  ..........  ...T....A.  .A........  ..........  .........T
P3              ..........  ..........  ...T....A.  .A........  ..........  ........TT
Ptv4            ..........  ..........  ...T....?.  .A........  ..........  ........TT
Ptv5            ..........  ..........  ..........  ..........  ..........  ..........
Ptv15           ..........  ..........  ..........  .........G  ..........  ..........
Tobar           ?.........  ..........  ...T....A.  .A........  ..........  ........TT
ANDERSON        .....A.T..  .G........  .CT.....A.  .AC....C..  CTG.......  ..A.C.....
```

```
            16262       16272       16282       16292       16301       16311
            |           |           |           |           |           |
A-176       CTCCCCCACC  CCAATATCAA  CAGACCTATC  TCCCC-TTAA  CAGTACATAG  TACATACATC
Al2         .......... ..G....... ..A....... CT.T...... .......... .........A.
Be1         .......... ..G....... ..A....... CT.T...... .......... .........A.
California  .......... ..G...C... ..A......A C..T...... .......... C........A.
DH1         .......... ..G...C... .......G.A CT........ .......... .........A.
Evered      .......... ..G...C... ..A......A C..T...... .......... C........A.
Frodo       .........T .......... ..A....... CT.T...... .......... .........A.
George      .......... ..G...C... .........A CT.T...... .......... .........A.
Georgette   .......... ..G...C... .......G.A CT........ .......... .........A.
Gigi        .......... ..G...C... ..A......A CT.T...... .......... C........A.
Gimble      .......... ..G...C... .........CA CT........ .......... .........A.
JF1         .......... .TG...C... ..A......A CT.T...... .......... C........A.
Kidevu      .........T .......... ..A....... CT.T...... .......... .........A.
Maggie      .......... ..G...C... ..A......A CT.TT..... .......... .........A.
Max         .......... .TG...C... ..A......A CT.T...... .......... C........A.
Me1         .......... ..G....... ..A....... CT.T...... .......... .........A.
Noel        .......... .......... ..A....... CT.T...... .......... .........A.
Pal         ........T. ..G....... ..A....... CT........ .......... .........A.
Prof        .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts2        .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts6        .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts12       .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts20       .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts21       .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts29       .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts35       .......... ..G...C... .........A CT.T...... .......... .........A.
Pts43       .......... ..GT..C... ..A....G.A CT......?? .?........ .........A?
Pts49       .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts50       .......... .TG...C... ..A......A CT.T...... .......... C........A.
Pts51       .......... .TG...C... ..A......A CT.T...... .......... C........A.
Ruhara      .......... ..G...C... .........A CT.T...... .......... .........A.
Sandi       .......... ..G....... ..A....... CT.T...... .......... .........A.
Skosha      .......... ..G...C... ..A......A CT.T...... .......... C........A.
Spindle     .......... ..G...C... ..A......A CT.T...... .......... C........A.
Sy2         .......... ..G....... ..A....... CT.T...... .......... .........A.
TT3         .......... ..G...C... ..A......A C..T...... .......... C........A.
Tubi        .......... ..G...C... ..A......A CT.T...... .......... C........A.
WL3         .......... ...G...... ..A....... CT........ .......... .........A.

A-60        .......... ..G...C... ..A.....CA CT........ .......... C........A.
A-175       .......... ..G...C... ..A.....CA CT........ .......... C........A.
C1          .......... ..G...C... ..A.....CA CT........ .......... C........A.
Berthe      ...T...... ..G...C... ..A......A C.TTT..... .......... .........A.
Gemini      .......... ..G...C... ..A.....CA CT........ .......... C........A.
Ivindo      .......... ..G...C... ..A.....CA CT........ .......... C........A.
Makokou     .......... ..G...C... ..A.....CA CT........ .......... C........A.
Milla       ..??...... ..G...C... ..A......A C......... ...CG..... .........A.
Mopia       .......... ..G...C... ..A.....CA CT........ .......... C........A.
Mouilla     .......... ..G...C... ..A.....CA CT........ .......... .........A.
Ntebe       ....T..... .TG...C... ..A......A CT........ .......... .........A.
Ponia       .......... ..G...C... ..A..T..CA CT........ ...C...... C........A.
Ptt33       .......... ..G...C... ..A.....CA CT........ .......... C........A.
Ptt34       .......... ..G...C... ..A.....CA CT........ .......... C........A.
Ptt35       .......... .TG...C... ..A.....CA CT........ .......... C........A.
Ptt36       .......... ..G...C... ..A......A C......... .......... .........A.
Ptt37       .......T.. ..G...C... .......GCA C......... T......... C........A.
Ptt43       ....T..... .TG...C... ..A......A CT........ .......... .........A.
Ptt59       .......... ..G...C... ..A.....CA CT........ .......... C........A.
Ptt62       .......... ..G...C... ..A.....CA CTT....... .......... C........A.
```

```
          16262      16272      16282      16292      16301      16311
          |          |          |          |          |          |
A-26      CTCCCCCACC CCGATATCAA CAGACCTACC CCCCC-TTGA CAGAACATAG TACATACAAT
A-33      .......... .TA...C... ..A....... .T........ ..A......A ..........
A-42      .......... ......C... ..A....... .......... ..A....... ..........
A-50      .......... .......... .......... ....-..... .......... ..........
A-52      .......... .......... .......... ....-..... ...?...... ..........
A-56      .......... .TA...C... ..A....... .T........ ..A......A ..........
A-62      .......... ......C... ..A...C... .......... ..AG...... ..........
A-69      .......... .......... .......... .......... .......... ..........
A-89      .......... ......C... ..A....... .......... ..A....... ..........
A-90      .......... .TA...C... ..A....... .T........ ..A......A ..........
A-94      .......... ......C... .......... T......... .......... .........C
A-97      .......... .TA...C... ..A....... .T........ ..A......A ..........
A-101     .......... ......C... ..A....... ....-..... ..A....... ..........
A-102     .......... ......C... .......... T......... .......... .........C
A-108     .......... .......... .......... .......... .......... ..........
A-118     .......... ......C... ..A....... .......... ..A....... ..........
A-128     .......... ......C... .......... T......... .......... .........C
A-129     .......... ......C... ..A....... .......... ..A....... ..........
A-130     .......... ......C... .......... T......... .......... .........C
A-131     .......... ......C... ..A....... .....C.... ..A....... ..........
A-136     .......... ......C... ..A....... .......... ..A....... ..........
A-137     .......... .TA...C... ..A....... .T........ ..A......A ..........
A-139     .......... .TA...C... ..A....... .T........ ..A......A ..........
A-172     .......... ......C... ..A.....T. ..TT....A. ..AG...... C.......G.
A-179     .......... .......... .......... .......... .......... ..........
A-182     .......... ......C... ..A...?... .......... ..AG...... ..........
A-192     .......... ......C... ..A.....T. ..TT....A. ..AG...... C.......G.
A-197     .......... ......C... ..A.....T. ..TT....A. ..AG...... C.......G.
A-199     .......... .......... .......... ....-..... .......... ..........
A-208     .......... ......C... .......... T......... ...?...... .........C
A-230     .......... ......C... ..A....... .......... ..A....... ..........
A-231     .......... ......C... .......... T......... .......... .........C
A-235     .......... ......C... ..A....... .T..-..... .......... ..........
A-239     .......... .......... .......... ....-..... .......... C.........
A-242     .......... .......... .......... .......... .......... ..........
A-268     .......... .......... ..A....... .......... .......... ..........
A-281     .......... ......C... ..A.....T. ..TT....A. ..AG...... C.......G.
A-282     .......... ......C... ..A....... ....-..... ..A....... ..........
A-283     .......... .......... .......... .......... .......... ..........
A-284     .......... .......... .......... .......... ...?...... ..........
A-285     .......... ......C... .......... T......... .......... .........C
A-286     .......... .TA...C... ..A....... .T........ ..A......A ..........
A-288     .......... .TA...C... ..A....... .T........ ..A......A ..........
A-290     .......... .......... ..A....... .......... ...?...... ..........
A-291     .......... ......C... ..A....... .......... ..A....... ..........
A-292     .......... .......... .......... .......... .......... ..........
C2        .......... ......C... ..A....... .......... ..A....... ..........
C3        .......... ......C... ........T. T......... .......... .........C
Edgar     .......... .......... .......... .......... ?......... ..........
NC588     ...?...... .TA...C... ..A....... .T........ ..A......A ..........
P1        ??????.... .......... .......... ......C... .......... ..........
P2a       .......... ......C... ..A....... ....-..... ..A....... ..........
P3        .......... ......C... ..A....... .......... T.A....... ..........
Ptv4      .......... ......C... ..A....... .......... ..A....... ..........
Ptv5      .......... .......... .......... ....-..... .......... ..........
Ptv15     .......... .......... .......... ..T...C... .......... ..........
Tobar     .?.??..... ......C... ..A....... .......... ..A....... ..........
ANDERSON  ...A.....T AG....C... ..A....... .A......A. ...T...... ......A.GC
```

```
            16321       16331       16341       16351       16361       16371
            |           |           |           |           |           |
A-176       CGTACATCAT  ACATAGCACA  TTACAGTTAA  ACCCATTCTC  GTCCCCACGG  ATGCCCCCCC
Al2         ......C...  ..........  .......C..  .T....C...  .C........  ..........
Be1         ......C...  ..........  .......C..  .T....C...  .C........  ..........
California  ......C...  ..........  .......C..  .T....C...  .C.....?..  ..........
DH1         ......C...  ..........  .......C..  .T....C...  .C........  ..........
Evered      ......C...  ..........  .......C..  .T....C...  .C........  ..........
Frodo       ......C...  ..........  ...T...C..  .T....C...  .C........  ..........
George      ......C...  ..........  .......C..  .T....C...  .C........  ..........
Georgette   ......C...  ..........  .......C..  .T....C...  .C........  ..........
Gigi        ..........  ..........  .......C..  .T........  .C........  ....??..??
Gimble      ......C...  ..........  .......C..  .T....C...  .C........  ..........
JF1         ......C...  ..........  .......C..  .T....C...  .C........  ..........
Kidevu      ......C...  ..........  ...T...C..  .T....C...  .C........  ....??????
Maggie      ......C...  ..........  .......C..  .T........  .C........  ..........
Max         ......C...  ..........  .......C..  .T....C...  .C........  ..........
Me1         ......C...  ..........  .......C..  .T....C...  .C........  ..........
Noel        ......C...  ..........  ...T...C..  .T....C...  .C........  .........T
Pal         ......C...  ..........  .......C..  .T....C...  .C........  ..........
Prof        ......C...  ..........  .......C..  .T....C...  .C........  ..........
Pts2        ......C...  ..........  .......C..  .T....C...  .C........  ..........
Pts6        ......C...  ..........  .......C..  .T....C..G  .G.??.....  ...???..??
Pts12       ......C...  ..........  .......C..  .T....C...  .C........  ....??????
Pts20       ......C...  ..........  .......C..  .T....C...  .C?.......  .....?????
Pts21       ......C...  ..........  .......C..  .T....C...  .C........  ....??????
Pts29       ......C...  ..........  .......C..  .T....C...  .C........  ..........
Pts35       ......C...  ..........  .......C..  .T....C...  .C........  ..........
Pts43       ......C?..  ..........  .......C..  .T....C...  .G........  ..........
Pts49       ......C...  ..........  .......C..  .T....C...  .C........  ..........
Pts50       ......C...  ..........  .......C..  .T....C...  .C........  ..........
Pts51       ......C...  ..........  .......C..  .T....C...  .C........  ..........
Ruhara      ......C...  ..........  .......C..  .T....C...  .C........  ..........
Sandi       ......C...  ..........  .......C..  .T....C...  .C........  ..........
Skosha      ..........  ..........  .......C..  .T........  .C........  ..........
Spindle     ..........  ..........  .......C..  .T........  .C........  ..........
Sy2         ......C...  ..........  .......C..  .T....C...  .C........  ..........
TT3         ......C...  ..........  .......C..  .T....C...  .C.???....  ..........
Tubi        ..........  ..........  .......C..  .T........  .C........  ..........
WL3         ......C...  ..........  ...T...C..  .T....C...  .C........  ..........

A-60        ..C...C..G  ..........  ...T...C..  .T........  ..........  ..........
A-175       ..C...C..G  ..........  ...T...C..  .T........  ..........  ..........
C1          ..C...C..G  ..........  .......C..  .T........  ..........  ..........
Berthe      ........GC  ..........  .......C..  .T....C..T  ..........  ..........
Gemini      ..C...C..G  ..........  ...T...C..  .T........  ..........  ..........
Ivindo      ..C...C..G  ..........  .......C..  .T........  ..........  ..........
Makokou     ..C...C..G  ..........  .......C..  .T........  ..........  ..........
Milla       ........G.  ..........  .CG....C..  .T....C..T  ..........  ..........
Mopia       ..C...C..G  ....G.....  .......C..  .T........  .C........  ..........
Mouilla     ..C...C..G  ..........  ...T...C..  .T........  ....T.....  ..........
Ntebe       ..C......C  ..........  .......C..  .T..G.C...  .C........  ..........
Ponia       ..C...C..G  ....G.....  .......C..  .T....C..T  .C........  ..........
Ptt33       ..C...C..G  ..........  ...T...C..  .T........  ..........  ..........
Ptt34       ..C...C..G  ..........  ...T...C..  .T........  ..........  ..........
Ptt35       ..C...C..G  ..........  .......C..  .T........  ..........  ..........
Ptt36       ........GC  ..........  .......C..  .T....C..T  ..........  ..........
Ptt37       ..C...C..G  ....G.....  .......C..  ..........  ..........  ..........
Ptt43       ..C......C  ..........  .......C..  .T..G.C...  .C........  .........?
Ptt59       ..C...C..G  ..........  ..GT...C..  .T........  ..........  .....?????
Ptt62       ..C...C..G  ..........  .......C..  .T........  ..........  ..........
```

```
          16321      16331      16341      16351      16361      16371
          |          |          |          |          |          |
A-26      CATACACCGT ACATAGCACA TTACAGTCAA ATTCTTCCTC GTCCCCACGG ATGCCCCCCC
A-33      ..C....... .......... .......... ..C.A....T .C........ ...AT.....
A-42      ..C....... .......... .......... ..C.A..... .C........ ...AT.....
A-50      .......... .......... .......... .C........ .......... ..........
A-52      .......... .......... .......... .C........ .......... ..........
A-56      ..C...T... .......... .......... ..C.A....T .C........ ...AT.....
A-62      ......T... .......... .......... ..C.A....T .......... ...A......
A-69      .......... .......... .......... .......... .......... ..........
A-89      ..C....... .......... .......... ..C.A..... .C........ ...AT.....
A-90      ..C....... .......... .......... ..C.A....T .C........ ...AT.....
A-94      .......... .......... .......... .CC.C..... .C........ ....T.....
A-97      ..C....... .......... .......... ..C.A....T .C........ ...AT.....
A-101     ..?....... .......... .......T.. ..C.A....T .C........ ...AT.....
A-102     .......... .......... .......... .CC.C.T... .C........ ....T.....
A-108     .......... .......... .......... .C........ .......... ..........
A-118     ..C....... .......... .......... ..C.A..... .C........ ...AT.....
A-128     .......... .......... .......... .CC.C..... .C........ ....T.....
A-129     ..C....... .......... .......T.. ..C.A....T .C........ ...AT.....
A-130     .......... .......... .......... .CC.C..... .C........ ....T.....
A-131     ..C....... .......... .......... ..C.A..... .C........ ...AT.....
A-136     ..C....... .......... .......... ..C.A..... .?........ ...AT.....
A-137     ..C....... .......... .......... ..C.A....T .C........ ...AT.....
A-139     ..C....... .......... .......... ..C.A....T .C........ ...AT.....
A-172     ..C.T..... .......... .......... ..C.A..... .......... ..........
A-179     .......... .......... .......... .......... .......... ..........
A-182     ......T... .......... .......... ..C.A....T .......... ...A?.....
A-192     ..C.T..... .......... .......... ..C.A..... .......... ..........
A-197     ..C.T..... .......... .......... ..C.A..... .......... ..........
A-199     .......... .......... .......... .C........ .......... ..........
A-208     .......... .......... .......... .CC.C..... .C........ ....T.....
A-230     ..C....... .......... .......... ..C.A..... .C........ ...AT.....
A-231     .......... .......... .......... .CC.C..... .......... ....T....T
A-235     ..C....... .......... .......T.. ..C.A....T .C........ ...AT.....
A-239     .......... .......... .......... .C........ .......... ..........
A-242     .......... .......... .......?.. ..?....... .......... ..........
A-268     .......... .......... .......... .C........ .......... ..........
A-281     ..C.T..... .......... .......... ..C.A..... .......... ..........
A-282     .......... .......... .......T.. ..C.A....T .C........ ...AT.....
A-283     .......... .......... .......?.. .......... .......... ..........
A-284     .......... .......... .......... ....?..... .......... ..........
A-285     .......... .......... .......... .CC.C..... .C........ ....T.....
A-286     ..C....... .......... .......... ..C.A....T .C........ ...AT.....
A-288     ..C...T... .......... .......... ..C.A....T .C........ ...AT.....
A-290     .......... .......... .......... .C........ .......... ..........
A-291     ..C....... .......... .......... ..C.A..... .C........ ...AT.....
A-292     .......... .......... .......... .......... .......... ..........
C2        ..C....... .......... .......... ..C.A..... .C........ ...AT.....
C3        .......... .......... .......... .CC.C..... .C........ ....T.....
Edgar     .......... .......... .......... .......... ?.....?.?? ??????????
NC588     T.C....... .......... .......... ..C.A....T .C........ ...AT.....
P1        .......... .......... .......... .......... .......... ..........
P2a       ..C....... .......... .......T.. ..C.A....T .C........ ...AT..???
P3        ..C....... .......... .......... ..C.A..... .C........ ...AT..???
Ptv4      ..C....... .......... .......... ..C.A..... .C........ ...AT.....
Ptv5      .......... .......... .......... .......... .......... ..........
Ptv15     .......... .......... .......... .C........ .......... ..........
Tobar     ..C....... .......... .......... ..C.A..... .C........ ...AT.....
ANDERSON  ...TT..... .......... .......... ..C.C.T... .......T.. ...A......
```

```
            16381      16391
            |          |
A-176       TCAGATAGGG ATCCCTTGTT
Al2         .......... G.????????
Be1         .......... G.????????
California  .......... ?.????????
DH1         ...?...... G.????????
Evered      .......... G?????????
Frodo       .......... G.????????
George      .......... G.????????
Georgette   .......... G.????????
Gigi        ?????G.... G.????????
Gimble      .......... G.????????
JF1         .......... G.????????
Kidevu      ?????????? ??????????
Maggie      .......... G.????????
Max         .......... G.????????
Me1         .......... G.????????
Noel        .......... G.????????
Pal         .......... G.????????
Prof        .......... G.????????
Pts2        .......... G.????????
Pts6        ?????????? ??????????
Pts12       ?????????? ??????????
Pts20       ?????????? ??????????
Pts21       ?????????? ??????????
Pts29       .......... G.????????
Pts35       .......... G.????????
Pts43       .......... G.????????
Pts49       .......... G.????????
Pts50       .......... G.????????
Pts51       .......... G.????????
Ruhara      .......... G.????????
Sandi       .......... G.????????
Skosha      .......... G.????????
Spindle     .......... G.????????
Sy2         .......... G.????????
TT3         .......... G.????????
Tubi        .......... G.????????
WL3         .......... G.????????

A-60        .........A ..........
A-175       .........A G.........
C1          .........A G.......C.
Berthe      .........A G?????????
Gemini      .........A G?????????
Ivindo      .........A G?????????
Makokou     .........A G.????????
Milla       .......... G.????????
Mopia       .........A G??????????
Mouilla     .........A G.????????
Ntebe       .......... G.????????
Ponia       .........A G?????????
Ptt33       .........A G.????????
Ptt34       .........A G.????????
Ptt35       ......G..A G.????????
Ptt36       .........A G??????????
Ptt37       .........A G.????????
Ptt43       ?????????? ??????????
Ptt59       ?????????? ??????????
Ptt62       ?????....A G.????????
```

```
            16381       16391
            |           |
A-26        TCAGATAGGA ATCCCTTGGT
A-33        ......G... ..........
A-42        .......... ........AC
A-50        .......... ..........
A-52        .......... ..........
A-56        ......G... ........AC
A-62        C.....G... ........AC
A-69        .......... ..........
A-89        .......... ........AC
A-90        ......G... ........AC
A-94        .......... ..........
A-97        ......G... ........AC
A-101       .......... ........AC
A-102       .......... ..........
A-108       .......... ..........
A-118       .......... ........AC
A-128       .......... ..........
A-129       .......... ........AC
A-130       .......... ..........
A-131       .......... ........AC
A-136       .......... ........AC
A-137       ......G... ........AC
A-139       ......G... ........AC
A-172       .......... ........AC
A-179       .......... ..........
A-182       .......... ........AC
A-192       .......... ........AC
A-197       .......... ........AC
A-199       .......... ..........
A-208       ........T. ..........
A-230       .......... ........AC
A-231       .......... ..........
A-235       .......... ........AC
A-239       .......... ..........
A-242       .......... ..........
A-268       .......... ..........
A-281       .......... ........AC
A-282       .?........ ........AC
A-283       .......... ..........
A-284       .......... ..........
A-285       ........T. ..........
A-286       ......G... ........AC
A-288       ......G... ........AC
A-290       .......... ..........
A-291       .......... ........AC
A-292       .......... ..........
C2          .......... ........AC
C3          .......... ..........
Edgar       ?????????? ??????????
NC588       ......G... ..????????
P1          .......... ..????????
P2a         ?????????? ??????????
P3          ?????????? ??????????
Ptv4        ...??????? ??????????
Ptv5        .......... ..????????
Ptv15       .......... ..????????
Tobar       .......... ..????????
ANDERSON    .........G G.......AC
```

APPENDIX B

Nucleotide sequence alignment of the mitochondrial NADH dehydrogenase subunit 2 (*ND2*) gene from humans, chimpanzees, bonobo and gorilla. Nucleotide positions are numbered according to Anderson *et al.* (1981). A dot (.) indicates sequence identity. Human sequences are from Africa (AFR), Europe (EUR), Asia (ASN) and Australia (AUS). Hs-ref is a previously published human sequence (Anderson *et al.* 1981). The majority of the chimpanzee sequences are from the *Pan troglodytes verus* subspecies, with the exception of A-175 and A-176 which are from *P. t. troglodytes* and/or *P. t. schweinfurthii* (see Section 3.4.1; Wise *et al.* 1997). Pt-ref is a previously published chimpanzee sequence (Horai *et al.* 1995). The bonobo and gorilla sequences are also published in Horai *et al.* (1995).

```
                4470        4480        4490        4500        4510
                |           |           |           |           |
Gorilla         ATTAACCCCC  TGGCCCAACC  CATCATCTAC  TCTACCATCT  TCGCAGGCAC
AFR1            .....T....  ..........  .G........  ..........  .T........
AFR2            .....T....  ..........  .G........  ..........  .T........
AFR3            .....T....  ..........  .G........  ..........  .T........
AFR4            .....T....  ..........  .G........  ..........  .T........
AFR5            .....T....  ..........  .G........  ..........  .T........
EUR1            .....T....  ..........  .G........  ..........  .T........
EUR2            .....T....  ..........  .G........  ..........  .T........
EUR3            .....T....  ..........  .G........  ..........  .T........
EUR4            .....T....  ..........  .G........  ..........  .T........
EUR5            .....T....  ..........  .G........  ..........  .T........
ASN1            .....T....  ..........  .G........  ..........  .T........
ASN2            .....T....  ..........  .G........  ..........  .T........
ASN3            .....T....  ..........  .G........  ..........  .T........
ASN4            .....T....  ..........  .G........  ..........  .T........
ASN5            .....T....  ..........  .G........  ..........  .T........
AUS1            .....T....  ..........  .G........  ..........  .T........
AUS2            .....T....  ..........  .G........  ..........  .T........
AUS3            .....T....  ..........  .G........  ..........  .T........
AUS4            .....T....  ..........  .G........  ..........  .T........
AUS5            .....T....  ..........  .G........  ..........  .T........
Hs-ref          .....T....  ..........  .G........  ..........  .T........
Bonobo          .....T....  .A........  ..........  ..........  ..........
A-175           .....T....  .A........  ..........  ..........  .T........
A-176           .....T....  .A........  ..........  ..........  .T........
A-281           .....T....  .A........  ..........  ..........  .TA.......
A-208           .....T....  .A........  ..........  ..........  .TA.......
A-231           .....T....  .A........  ..........  ..........  .TA.......
A-285           .....T....  .A........  ..........  ..........  .TA.......
A-199           .....T....  .A........  ..........  ..........  .TA.......
A-241           .....T....  .A........  ..........  ..........  .TA.......
A-292           .....T....  .A........  ..........  ..........  .TA.......
A-50            .....T....  .A........  ..........  ..........  .TA.......
A-179           .....T....  .A........  ..........  ..........  .TA.......
A-283           .....T....  .A........  ..........  ..........  .TA.......
A-290           .....T....  .A........  ..........  ..........  .TA.......
A-242           .....T....  .A........  ..........  ..........  .TA.......
A-182           .....T....  .A........  ..........  ..........  .TA.......
A-288           .....T....  .A........  ..........  .........C  .TA.......
Pt-ref          .....T....  .A........  ..........  .........C  .TA.......
A-291           ..........  .A........  ..........  ..........  .TA.......
A-282           ..........  .A........  ..........  ..........  .TA.......
A-287           ..........  .A........  ..........  ..........  .TA.......
A-235           ..........  .A........  ..........  ..........  .TA.......
```

```
                4520        4530        4540        4550        4560
                |           |           |           |           |
Gorilla         TCTTATTACA  GCACTAAGCT  CCCACTGATT  TTTTGCCTGA  GTAGGCCTAG
AFR1            A..C..C...  ..G.......  .G........  ....A.....  ..........
AFR2            A..C..C...  ..G.......  .G........  ....A.....  ..........
AFR3            A..C..C...  ..G.......  .G........  ....A.....  ..........
AFR4            A..C..C...  ..G.......  .G........  ....A.....  ..........
AFR5            A..C..C...  ..G.......  .G........  ....A.....  ..........
EUR1            A..C..C...  ..G.......  .G........  ....A.....  ..........
EUR2            A..C..C...  ..G.......  .G........  ....A.....  ..........
EUR3            A..C..C...  ..G.......  .G........  ....A.....  ..........
EUR4            A..C..C...  ..G.......  .G........  ....A.....  ..........
EUR5            A..C..C...  ..G.......  .G........  ....A.....  ..........
ASN1            A..C..C...  ..G.......  .G........  ....A.....  ..........
ASN2            A..C..C...  ..G.......  .G........  ....A.....  ..........
ASN3            A..C..C...  ..G.......  .G........  ....A.....  ..........
ASN4            A..C..C...  ..G.......  .G........  ....A.....  ..........
ASN5            A..C..C...  ..G.......  .G........  ....A.....  ..........
AUS1            A..C..C...  ..G.......  .G........  ....A.....  ..........
AUS2            A..C..C...  ..G.......  .G........  ....A.....  ..........
AUS3            A..C..C...  ..G.......  .G........  ....A.....  ..........
AUS4            A..C..C...  ..G.......  .G........  ....A.....  ..........
AUS5            A..C..C...  ..G.......  .G........  ....A.....  ..........
Hs-ref          A..C..C...  ..G.......  .G........  ....A.....  ..........
Bonobo          AT.C......  .TG.......  .A........  ...CA.....  ..........
A-175           A..C......  ..G.......  .A........  ...CA.....  ..........
A-176           A..C.....G  ..G.......  .A........  ...CA.....  ..........
A-281           G..C......  ..G.......  .A........  ...CA.....  ..........
A-208           G..C......  ..G.......  .A........  ...CA.....  ..........
A-231           G..C......  ..G.......  .A........  ...CA.....  ..........
A-285           G..C......  ..G.......  .A........  ...CA.....  ..........
A-199           G..C......  ..G.......  .A........  ...CA.....  ..........
A-241           G..C......  ..G.......  .A........  ...CA.....  ..........
A-292           G..C......  ..G.......  .A........  ...CA.....  ..........
A-50            G..C......  ..G.......  .A........  ...CA.....  ..........
A-179           G..C......  ..G.......  .A........  ...CA.....  ..........
A-283           G..C......  ..G.......  .A........  ...CA.....  ..........
A-290           G..C......  ..G.......  .A........  ...CA.....  ..........
A-242           G..C......  ..G.......  .A........  ...CA.....  ..........
A-182           G..C......  ..G.......  .A........  ...CA.....  ..........
A-288           G..C......  ..G.......  .A........  ...CA.....  ..........
Pt-ref          G..C......  ..G.......  .A........  ...CA.....  ..........
A-291           G..C......  ..G.......  .A........  ...CA.....  ..........
A-282           G..C......  ..G.......  .A........  ...CA.....  ..........
A-287           G..C......  ..G.......  .A........  ...CA.....  ..........
A-235           G..C......  ..G.......  .A........  ...CA.....  ..........
```

```
                4570        4580        4590        4600        4610
                |           |           |           |           |
Gorilla         AAATAAACAT  ACTAGCTTTT  ATCCCAGTCC  TAACCAAAAA  AATAAATCCC
AFR1            .......... G......... ..T.....T. .......... ......C..T
AFR2            .......... G......... ..T.....T. .......... ......C..T
AFR3            .......... G......... ..T.....T. .......... ......C..T
AFR4            .......... G......... ..T.....T. .......... ......C..T
AFR5            .......... G......... ..T.....T. .......... ......C..T
EUR1            .......... G......... ..T.....T. .......... ......C..T
EUR2            .......... G......... ..T.....T. .......... ......C..T
EUR3            .......... G......... ..T.....T. .......... ......C..T
EUR4            .......... G......... ..T.....T. .......... ......C..T
EUR5            .......... G......... ..T.....T. .......... ......C..T
ASN1            .......... G......... ..T.....T. .......... ......C..T
ASN2            .......... G......... ..T.....T. .......... ......C..T
ASN3            .......... G......... ..T.....T. .......... ......C..T
ASN4            .......... G......... ..T.....T. .......... ......C..T
ASN5            .......... G......... ..T.....T. .......... ......C..T
AUS1            .......... G......... ..T.....T. .......... ......C..T
AUS2            .......... G......... ..T.....T. .......... ......C..T
AUS3            .......... G......... ..T.....T. .......... ......C..T
AUS4            .......... G......... ..T.....T. .......... ......C..T
AUS5            .......... G......... ..T.....T. .......... ......C..T
Hs-ref          .......... G......... ..T.....T. .......... ......C..T
Bonobo          .......T.. .......... .......... .......... .....GC...
A-175           .......T.. .......... ......A... .......... .....GC...
A-176           .......T.. .......... ......A... .......... .....GC...
A-281           .......T.. .......... ......A... .......... .....GC...
A-208           .......T.. .......... ......A... .......... .....GC...
A-231           .......T.. .......... ......A... .......... .....GC...
A-285           .......T.. .......... ......A... .......... .....GC...
A-199           .......T.. .......... ......A... .......... .....GC...
A-241           .......T.. .......... ......A... .......... .....GC...
A-292           .......T.. .......... ......A... .......... .....GC...
A-50            .......T.. .......... ......A... .......... .....GC...
A-179           .......T.. .......... ......A... .......... .....GC...
A-283           .......T.. .......... ......A... .......... .....GC...
A-290           .......T.. .......... ......A... .......... .....GC...
A-242           .......T.. .......... ......A... .......... .....GC...
A-182           .......T.. .......... ......A... .......... .....GC...
A-288           .......T.. .......... ......A... .......... .....GC...
Pt-ref          .......T.. .......... ......A... .......... .....GC...
A-291           .......T.. .......... ......A... .......... .....GC...
A-282           .......T.. .......... ......A... .......... .....GC...
A-287           .......T.. .......... ......A... .......... .....GC...
A-235           .......T.. .......... ......A... .......... .....GC...
```

```
                4620        4630        4640        4650        4660
                |           |           |           |           |
Gorilla         CGCTCCACAG  AAGCCGCCAT  CAAATATTTC  CTCACACAAG  CAACTGCATC
AFR1            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AFR2            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AFR3            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AFR4            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AFR5            ..T.......  ....T.....  ...G......  .....G....  ....C.....
EUR1            ..T.......  ....T.....  ...G......  .....G....  ....C.....
EUR2            ..T.......  ....T.....  ...G......  .....G....  ....C.....
EUR3            ..T.......  ....T.....  ...G......  .....G....  ....C.....
EUR4            ..T.......  ....T.....  ...G......  .....G....  ....C.....
EUR5            ..T.......  ....T.....  ...G......  .....G....  ....C.....
ASN1            ..T.......  ....T.....  ...G......  .....G....  ....C.....
ASN2            ..T.......  ....T.....  ...G......  .....G....  ....C.....
ASN3            ..T.......  ....T.....  ...G......  .....G....  ....C.....
ASN4            ..T.......  ....T.....  ...G......  .....G....  ....C.....
ASN5            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AUS1            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AUS2            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AUS3            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AUS4            ..T.......  ....T.....  ...G......  .....G....  ....C.....
AUS5            ..T.......  ....T.....  ...G......  .....G....  ....C.....
Hs-ref          ..T.......  ....T.....  ...G......  .....G....  ....C.....
Bonobo          ..........  ....T.....  ......C..T  ..........  ....C..G..
A-175           ..........  ....T.....  ......C..T  ..........  .......G..
A-176           ..........  ....T.....  ......C..T  ..........  .......G..
A-281           ..........  ..........  ......C..T  ..........  .......G..
A-208           ..........  .........C  ......C..T  ..........  .......G..
A-231           ..........  .........C  ......C..T  ..........  .......G..
A-285           ..........  .........C  ......C..T  ..........  .......G..
A-199           ..........  .........C  ......C..T  ..........  .......G..
A-241           ..........  .........C  ......C..T  ..........  .......G..
A-292           ..........  .........C  ......C..T  ..........  .......G..
A-50            ..........  .........C  ......C..T  ..........  .......G..
A-179           ..........  .........C  ......C..T  ..........  ..........
A-283           ..........  .........C  ......C..T  ..........  ..........
A-290           ..........  .........C  ......C..T  ..........  .......G..
A-242           ..........  .........C  ......C..T  ..........  .......G..
A-182           ..........  ..........  ......C..T  ..........  .......G..
A-288           ..........  ..........  ......C..T  ..........  .......G..
Pt-ref          ..........  ..........  ......C..T  ..........  .......G..
A-291           ..........  ..........  ......C..T  ..........  .......G..
A-282           ..........  ..........  ......C..T  ..........  .......G..
A-287           ..........  ..........  ......C..T  ..........  .......G..
A-235           ..........  ..........  ......C..T  ..........  .......G..
```

	4670	4680	4690	4700	4710
Gorilla	CATAATCCTC	CTAATAGCCA	TCCTCTCCAA	CAACATACTC	TCCGGACAAT
AFR1	.....T	.....T.	.....T...	...T.....	.....
AFR2	.....T	.....T.	.....T...	...T.....	.....
AFR3	.....T	.....T.	.....T...	...T.....	.....
AFR4	.....T	.....T.	.....T...	...T.....	.....
AFR5	.....T	.....T.	.....T...	...T.....	.....
EUR1	.....T	.....T.	.....T...	...T.....	.....
EUR2	.....T	.....T.	.....T...	...T.....	.....
EUR3	.....T	.....T.	.....T...	...T.....	.....
EUR4	.....T	.....T.	.....T...	...T.....	.....
EUR5	....G....T	.....T.	.....T...	...T.....	.....
ASN1	.....T	.....T.	.....T...	...T.....	.....
ASN2	.....T	.....T.	.....T...	...T.....	.....
ASN3	.....T	.....T.	.....T...	...T.....	.....
ASN4	.....T	.....T.	.....T...	...T.....	.....
ASN5	.....T	.....T.	.....T...	...T.....	.....
AUS1	.....T	.....T.	.....T...	...T.....	.....
AUS2	.....T	.....T.	.....T...	...T.....	.....
AUS3	.....T	.....T.	.....T...	...T.....	.....
AUS4	.....T	.....T.	.....T...	...T.....	.....
AUS5	.....T	.....T.	.....T...	...T.....	.....
Hs-ref	.....T	.....T.	.....T...	...T.....	.....
Bonobo	...G..T...	.....T.	.....	.....	.....
A-175	.....T...	..G.....T.	.....	..G.....	.....
A-176	.....T...	..G.....T.	.....	..G.....	.....
A-281	.....T...	..G.....T.	.....	..G.....	.....
A-208	.....T...	..G.....T.	.....	..G.....	.....
A-231	.....T...	..G.....T.	.....	..G.....	.....
A-285	.....T...	..G.....T.	.....	..G.....	.....
A-199	.....T...	..G.....T.	.....	..G.....	.....
A-241	.....T...	..G.....T.	.....	..G.....	.....
A-292	.....T...	..G.....T.	.....	..G.....	.....
A-50	.....T...	..G.....T.	.....	..G.....	.....
A-179	.....T...	..G.....T.	.....	..G.....	.....
A-283	.....T...	..G.....T.	.....	..G.....	.....
A-290	.....T...	..G.....T.	.....	..G.....	.....
A-242	.....T...	..G.....T.	.....	..G.....	.....
A-182	.....T...	..G.....T.	.....	..G.....	.....
A-288	.....T...	..G.....T.	.....	..G.....	.....
Pt-ref	.....T...	..G.....T.	.....	..G.....	.....
A-291	.....T...	..G.....T.	.....	..G.....	.....
A-282	.....T...	..G.....T.	.....	..G.....	.....
A-287	.....T...	..G.....T.	.....	..G.....	.....
A-235	.....T...	..G.....T.	.....	..G.....	.....

```
                4720        4730        4740        4750        4760
                |           |           |           |           |
Gorilla         GAACCACAAC  CAATGCCACT  AATCAATACT  CATCATTAAT  GATCGTAGTA
AFR1            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AFR2            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AFR3            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AFR4            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AFR5            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
EUR1            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
EUR2            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
EUR3            ......T...  ..G.A.T..C  ..........  ..........  A...A..A.G
EUR4            ......T...  ....A.T..C  ..........  ..........  A...A..A..
EUR5            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
ASN1            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
ASN2            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
ASN3            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
ASN4            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
ASN5            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AUS1            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AUS2            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AUS3            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AUS4            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
AUS5            ......T...  ....A.T..C  ..........  ..........  A...A..A.G
Hs-ref          ......T...  ....A.T..C  ..........  ..........  A...A..A..
Bonobo          ......T...  ...CA.T..C  ..........  ..........  A..TA..AC.
A-175           ......T..T  ....A.T.TC  ..........  ..........  A..TA..A..
A-176           ......T..T  ....A.T.TC  ..........  ..........  A..TA..A..
A-281           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-208           ......T...  ....A.T..C  .........C  ..........  A..TA..A..
A-231           ......T...  ....A.T..C  .........C  ..........  A..TA..A..
A-285           ......T...  ....A.T..C  .........C  ..........  A..TA..A..
A-199           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-241           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-292           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-50            ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-179           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-283           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-290           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-242           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-182           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-288           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
Pt-ref          ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-291           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-282           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-287           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
A-235           ......T...  ....A.T..C  ..........  ..........  A..TA..A..
```

```
                4770        4780        4790        4800        4810
                |           |           |           |           |
Gorilla         GCTATAGCTA  TAAAACTAGG  AATAGCCCCC  TTTCACTTCT  GAGTGCCAGA
AFR1            .........A  ..........  ..........  ..........  ....C.....
AFR2            .........A  ..........  ..........  ..........  ....C.....
AFR3            .........A  ..........  ..........  ..........  ....C.....
AFR4            .........A  ..........  ..........  ..........  ....C.....
AFR5            .........A  ..........  ..........  ..........  ....C.....
EUR1            .........A  ..........  ..........  ..........  ....C.....
EUR2            .........A  ..........  ..........  ..........  ....C.....
EUR3            .........A  ..........  ..........  ..........  ....C.....
EUR4            .........A  ..........  ..........  ..........  ....C.....
EUR5            .........A  ..........  ..........  ..........  ....C.....
ASN1            .........A  ..........  ..........  ..........  ....C.....
ASN2            .........A  ..........  ..........  ..........  ....C.....
ASN3            .........A  ..........  ..........  ..........  ....C.....
ASN4            .........A  ..........  ..........  ..........  ....C.....
ASN5            .........A  ..........  ..........  ..C.......  ....C.....
AUS1            .........A  ..........  ..........  ..........  ....C.....
AUS2            .........A  ..........  ..........  ..........  ....C.....
AUS3            .........A  ..........  ..........  ..........  ....C.....
AUS4            .........A  ..........  ..........  ..........  ....C.....
AUS5            .........A  ..........  ..........  ..........  ....C.....
Hs-ref          .........A  ..........  ..........  ..........  ....C.....
Bonobo          .....G..A.  ..........  T.........  ..........  ....T.....
A-175           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-176           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-281           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-208           ..A.....A.  ..........  ..........  ........T.  ....T.....
A-231           ..A.....A.  ..........  ..........  ........T.  ....T.....
A-285           ..A.....A.  ..........  ..........  ........T.  ..A.T.....
A-199           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-241           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-292           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-50            ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-179           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-283           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-290           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-242           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-182           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-288           ..A..G..A.  ..........  ..........  ........T.  ....T.....
Pt-ref          ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-291           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-282           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-287           ..A..G..A.  ..........  ..........  ........T.  ....T.....
A-235           ..A..G..A.  ..........  ..........  ........T.  ....T.....
```

	4820	4830	4840	4850	4860
Gorilla	AGTCACCCAA	GGCACCCCCC	TAATGTCTGG	CCTACTCCTC	CTCACATGAC
AFR1	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AFR2	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AFR3	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AFR4	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AFR5	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
EUR1	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
EUR2	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
EUR3	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
EUR4	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
EUR5	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
ASN1	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
ASN2	G..T.....	.....T.	.G.CA..C..	T..G..T..T	.....
ASN3	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
ASN4	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
ASN5	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AUS1	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AUS2	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AUS3	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AUS4	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
AUS5	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
Hs-ref	G..T.....	.....T.	.G.CA..C..	...G..T..T	.....
Bonobo	...T.....	.....	....A..C..	.....	.....
A-175	...T.....	.....T....	....A.....	.....	.....
A-176	...T.....	.....T....	....A.....	.....	.....
A-281	...T.....	.....	....A..C..	.....	.....
A-208	...T.....	.....	....A..C..	.....	.....
A-231	...T.....	.....	....A..C..	.....	.....
A-285	...T.....	.....	....A..C..	.....	.....
A-199	...T.....	.....	....A..C..	.....	.....
A-241	...T.....	.....	....A..C..	.....	.....
A-292	...T.....	.....	....A..C..	.....	.....
A-50	...T.....	.....	....A..C..	.....	.....
A-179	...T.....	.....	....A..C..	.....	.....
A-283	...T.....	.....	....A..C..	.....	.....
A-290	...T.....	.....	....A..C..	.....	.....
A-242	...T.....	.....	....A..C..	.....	.....
A-182	...T.....	.....	....A..C..	.....	.....
A-288	...T.....	.....	....A..C..	.....	.....
Pt-ref	...T.....	.....	....A..C..	.....	.....
A-291	...T.....	.....	....A..C..	...G.....	.....
A-282	...T.....	.....	....A..C..	.....	.....
A-287	...T.....	.....	....A..C..	.....	.....
A-235	...T.....	.....	....A..C..	.....	.....

```
                4870        4880        4890        4900        4910
                |           |           |           |           |
Gorilla         AAAAACTAGC  CCCTATGTCA  ATCATATACC  AAATTTCCTC  GTCAACAAAT
AFR1            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AFR2            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AFR3            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AFR4            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AFR5            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
EUR1            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
EUR2            ..........  ...C..C...  ..........  ....C..TC.  C...CT.G.C
EUR3            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
EUR4            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
EUR5            ..........  ...C..C...  ..........  ....C..TC.  C...CT.G.C
ASN1            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
ASN2            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
ASN3            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
ASN4            ..........  ......C...  ..........  ....C..TC.  C...CT...C
ASN5            ..........  ......C...  ..........  ....C..TC.  C...CT...C
AUS1            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AUS2            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AUS3            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AUS4            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
AUS5            ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
Hs-ref          ..........  ...C..C...  ..........  ....C..TC.  C...CT...C
Bonobo          ..........  ...C..C...  ..T.......  ....A.....  A...CTG..C
A-175           .....T....  ...C..T...  ..T.......  ....C.....  A...CTG..C
A-176           .....T....  ...C..T...  ..T.......  ....C.....  A...CTG..C
A-281           .....T....  ......T...  ..T.......  ....C.....  A...CTG..C
A-208           .....T....  ......T...  ..T.......  ....C.....  A...CTG..C
A-231           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-285           .....T....  ......C...  ..T.....T.  ....C.....  A...CTG..C
A-199           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-241           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-292           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-50            .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-179           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-283           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-290           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-242           .....T....  ......T...  ..T.....T.  ....C.....  A...CTG..C
A-182           .....T....  ......T...  ..T.......  ....C.....  A...CTG..C
A-288           .....T....  ......T...  ..T.......  ....C.....  A...CTG..C
Pt-ref          .....T....  ......T...  ..T.......  ....C.....  A...CTG..C
A-291           .....T....  ......T...  ..T.......  ....C.....  A...CTG..C
A-282           .....T....  ......T...  ..T.......  ....C...C.  A...CTG..C
A-287           .....T....  ......T...  ..T.......  ....C...C.  A...CTG..C
A-235           .....T....  ......T...  ..T.......  ....C.....  A...CTG..C
```

```
                4920        4930        4940        4950        4960
                |           |           |           |           |
Gorilla         GTAAGCCTTC  TCCTCACTCT  TTCAATCCTA  TCCATCCTAG  CAGGCAGCTG
AFR1            ..........  ..........  C......T..  ......A...  .......T..
AFR2            ..........  ..........  C......T..  ......A...  .......T..
AFR3            ..........  ..........  C......T..  ......A...  .......T..
AFR4            ..........  ..........  C......T..  ......A...  .......T..
AFR5            ..........  ..........  C......T..  ......A...  .......T..
EUR1            ..........  ..........  C......T..  ......A...  .......T..
EUR2            ..........  ..........  C......T..  ......A...  .......T..
EUR3            ..........  ..........  C......T..  ......A...  .......T..
EUR4            ..........  ..........  C......T..  ......A...  .......T..
EUR5            ..........  ..........  C......T..  ......A...  .......T..
ASN1            ..........  ..........  C......T..  ......A...  .......T..
ASN2            ..........  ..........  C......T..  ......A...  .......T..
ASN3            ..........  ..........  C......T..  ......A...  .......T..
ASN4            ..........  ..........  C......T..  ......A...  .......T..
ASN5            ..........  ..........  C......T..  ......A...  .......T..
AUS1            ..........  ..........  C......T..  ......A...  .......T..
AUS2            ..........  ..........  C......T..  ......A...  .......T..
AUS3            ..........  ..........  C......T..  ......A...  .......T..
AUS4            ..........  ..........  C......T..  ......A...  .......T..
AUS5            ..........  ..........  C......T..  ......A...  .......T..
Hs-ref          ..........  ..........  C......T..  ......A...  .......T..
Bonobo          ....A.....  .......C..  .......T.G  .....TA...  ..........
A-175           ..........  .......C..  .......T.G  .....TA...  ..........
A-176           ..........  .......C..  .......T.G  .....TA...  ..........
A-281           ....A.....  .......C..  .........G  .....TA...  ..........
A-208           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-231           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-285           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-199           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-241           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-292           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-50            ....A.....  .......C..  .......T.G  .....TA...  ..........
A-179           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-283           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-290           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-242           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-182           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-288           ....A.....  .......C..  .......T.G  .....TA...  ..........
Pt-ref          ....A.....  .......C..  .......T.G  .....TA...  ..........
A-291           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-282           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-287           ....A.....  .......C..  .......T.G  .....TA...  ..........
A-235           ....A.....  .......C..  .......T.G  .....TA...  ..........
```

	4970	4980	4990	5000	5010
Gorilla	AGGCGGACTA	AACCAAACTC	AACTACGCAA	GATTCTAGCA	TACTCCTCAA
AFR1	...T...T..	.....C.	.G.....	A..CT.....	.....
AFR2	...T...T..	.....C.	.G.....	A..CT.....	.....
AFR3	...T...T..	.....C.	.G.....	A..CT.....	.....
AFR4	...T...T..	.....C.	.G.....	A..CT.....	.....
AFR5	...T...T..	.....C.	.G.....	A..CT.....	.....
EUR1	...T...T..	.....C.	.G.....	A..CT.....	.....
EUR2	...T...T..	.....C.	.G.....	A..CT.....	.....
EUR3	...T...T..	.....C.	.G.....	A..CT.....	.....
EUR4	...T...T..	.....C.	.G.....	A..CT.....	.....
EUR5	...T...T..	.....C.	.G.....	A..CT.....	.....
ASN1	...T...T..	.....C.	.G.....	A..CT.....	.....
ASN2	...T...T..	.....C.	.G.....	A..CT.....	.....
ASN3	...T...T..	.....C.	.G.....	A..CT.....	.....
ASN4	...T...T..	.....C.	.G.....	A..CT.....	.....
ASN5	...T...T..	.....C.	.G.....	A..CT.....	.....
AUS1	...T...T..	.....C.	.G.....	A..CT.....	.....
AUS2	...T...T..	.....C.	.G.....	A..CT.....	.....
AUS3	...T...T..	.....C.	.G.....	A..CT.....	.....
AUS4	...T...T..	.....C.	.G.....	A..CT.....	.....
AUS5	...T...T..	.....C.	.G.....	A..CT.....	.....
Hs-ref	...T...T..	.....G..C.	.G.....	A..CT.....	.....
Bonobo	.....	.....C.	.....	A..C.....	.....
A-175	.....	.....C.	.....	A..C.....	.....
A-176	.....	.....C.	.....	A..C.....	.....
A-281	.....	.....C.	.....	A..C.....	.....
A-208	.....	.....C.	.....	A..C.....	.....
A-231	.....	.....C.	.....	A..C.....	.....
A-285	.....	.....C.	.....	A..C.....	.....
A-199	.....	.....C.	.....	A..C.....	.....
A-241	.....	.....C.	.....	A..C.....	.....
A-292	.....	.....C.	.....	A..C.....	.....
A-50	.....	.....C.	.....	A..C.....	.....
A-179	.....	.....C.	.....	A..C.....	.....
A-283	.....	.....C.	.....	A..C.....	.....
A-290	.....	.....C.	.....	A..C.....	.....
A-242	.....	.....C.	.....	A..C.....	.....
A-182	.....	.....C.	.....	A..C.....	.....
A-288	.....	.....C.	.....	A..C.....	.....
Pt-ref	.....	.....C.	.....	A..C.....	.....
A-291	.....	.....C.	.....	A..C.....	.....
A-282	.....	.....C.	.....	A..C.....	.....
A-287	.....	.....C.	.....	A..C.....	.....
A-235	.....	.....C.	.....	A..C.....	.....

	5020	5030	5040	5050	5060
Gorilla	TCACCCATGT	AGGATGAATA	ATAGCAGTTC	TACCATATAA	CCCTAATATA
AFR1	.T.....CA.	....	....	....G..C..	......C...
AFR2	.T.....CA.	....	....	....G..C..	......C...
AFR3	.T.....CA.	....	....	....G..C..	......C...
AFR4	.T.....CA.	....	....	....G..C..	......C...
AFR5	.T.....CA.	....	....	....G..C..	......C...
EUR1	.T.....CA.	....	....	....G..C..	......C...
EUR2	.T.....CA.	....	....	....G..C..	......C...
EUR3	.T.....CA.	....	....	....G..C..	......C...
EUR4	.T.....CA.	....	....	....G..C..	......C...
EUR5	.T.....CA.	....	....	....G..C..	......C...
ASN1	.T.....CA.	....	....	....G..C..	......C...
ASN2	.T.....CA.	....	....	....G..C..	......C...
ASN3	.T.....CA.	....	....	....G..C..	......C...
ASN4	.T.....CA.	....	....	....G..C..	......C...
ASN5	.T.....CA.	....	....	....G..C..	......C...
AUS1	.T.....CA.	....	....	....G..C..	......C...
AUS2	.T.....CA.	....	....	....G..C..	......C...
AUS3	.T.....CA.	....	....	....G..C..	......C...
AUS4	.T.....CA.	....	....	....G..C..	......C...
AUS5	.T......A.	....	....	....G..C..	......C...
Hs-ref	.T.....CA.	....	....	....G..C..	......C...
Bonobo	.......CA.	...C......	....	.......C..	......C...
A-175	.......CA.	...C......	....	.......C..	......C...
A-176	.......CA.	...C......	........C.	.......C..	......C...
A-281	.......CA.	...C......	........C.	....	......C...
A-208	.......CA.	...C......	........C.	....	......C...
A-231	.......CA.	...C......	........C.	....	......C...
A-285	.......CA.	...C......	........C.	....	......C...
A-199	.......CA.	...C......	........C.	....	...C..C...
A-241	.......CA.	...C......	........C.	....	...C..C...
A-292	.......CA.	...C......	........C.	....	...C..C...
A-50	.......CA.	...C......	........C.	....	...C..C...
A-179	.......CA.	...C......	........C.	....	...C..C...
A-283	.......CA.	...C......	........C.	....	...C..C...
A-290	.......CA.	...C......	........C.	....	...C..C...
A-242	.......CA.	...C......	........C.	....	...C..C...
A-182	........A.	...C......	........C.	....	......C...
A-288	.......CA.	...C......	........C.	....	......C...
Pt-ref	.......CA.	...C......	........C.	....	......C...
A-291	.......CA.	...C......	........C.	....	......C...
A-282	.......CA.	...C......	........C.	....	......C...
A-287	.......CA.	...C......	........C.	....	......C...
A-235	.......CA.	...C......	........C.	....	......C...

	5070	5080	5090	5100	5110
Gorilla	ACTATTCTTA	ATCTGACCAT	TTATATTATC	CTCACTACTA	CCACATTCCT
AFR1	..C.....	..T.A..T..	....	..A.....	..G.....
AFR2	..C.....	..T.A..T..	....	..A.....	..G.....
AFR3	..C.....	..T.A..T..	C.........	..A.....	..G.....
AFR4	..C.....	..T.A..T..	....	..A.....	..G.....
AFR5	..C.....	..T.A..T..	....	..A.....	..G.....
EUR1	..C.....	..T.A..T..	....	..A.....	..G.....
EUR2	..C.....	..T.A..T..	....	..A.....	..G.....
EUR3	..C.....	..T.A..T..	....	..A.....	..G.....
EUR4	..C.....	..T.A..T..	....	..A.....	..G.....
EUR5	..C.....	..T.A..T..	....	..A.....	..G.....
ASN1	..C.....	..T.A..T..	....	..A.....	..G.....
ASN2	..C.....	..T.A..T..	....	..A.....	..G.....
ASN3	..C.....	..T.A..T..	....	..A.....	..G.....
ASN4	..C.....	..T.A..T..	....	..A.....	..G.....
ASN5	..C.....	..T.A..T..	....	..A.....	..G.....
AUS1	..C.....	..T.A..T..	....	..A.....	..G.....
AUS2	..C.....	..T.A..T..	....	..A.....	..G.....
AUS3	..C.....	..T.A..T..	....	..A.....	..G.....
AUS4	..C.....	..T.A..T..	....	..A.....	..G.....
AUS5	..C.....	..T.A..T..	....	..A.....	..G.....
Hs-ref	..C.....	..T.A..T..	....	..A.....	..G.....
Bonobo	..C.....	..T.A..T..	......C...	..A.....	....
A-175	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-176	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-281	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-208	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-231	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-285	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-199	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-241	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-292	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-50	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-179	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-283	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-290	..C.....	..T.A..T..	...C..C...	..T.....	..G....T..
A-242	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..
A-182	..C.....	..T.A.....	...C..C...	..A.....	..G....T..
A-288	..C.....	..T.A.....	...C..C...	..A.....	..G....T..
Pt-ref	..C.....	..T.A.....	...C..C...	..A.....	..G....T..
A-291	..C.....	..T.A.....	...C..C...	..A.....	..G....T..
A-282	..C.....	..T.A.....	...C..C...	..A.....	..G....T..
A-287	..C.....	..T.A.....	...C..C...	..A.....	..G....T..
A-235	..C.....	..T.A..T..	...C..C...	..A.....	..G....T..

	5120	5130	5140	5150	5160
	|	|	|	|	|
Gorilla	ATTACTCAAC	CTAAGCTCCA	GCACCACAAC	CTTACTACTA	TCTCGTACTT
AFR1	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AFR2	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AFR3	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AFR4	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AFR5	.C.....	T...A.....	.....G..	.C.....	.....C..C.
EUR1	.C.....	T...A.....	.....G..	.C.....	.....C..C.
EUR2	.C.....	T...A.....	.....G..	.C.....	.....C..C.
EUR3	.C.....	T...A.....	.....G..	.C.....	.....C..C.
EUR4	.C.....	T...A.....	.....G..	.C.....	.....C..C.
EUR5	.C.....	T...A.....	.....	.C.....	.....C..C.
ASN1	.C.....	T...A.....	.....G..	.C.....	.....C..C.
ASN2	.C.....	T...A.....	.....G..	.C.....	.....C..C.
ASN3	.C.....	T...A.....	.....G..	.C.....	.....C..C.
ASN4	.C.....	T...A.....	.....G..	.C.....	.....C..C.
ASN5	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AUS1	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AUS2	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AUS3	.C.....	T...A.....	.....	.C.....	.....C..C.
AUS4	.C.....	T...A.....	.....G..	.C.....	.....C..C.
AUS5	.C.....	T...A.....	.....G..	.C.....	.....C..C.
Hs-ref	.C.....	T...A.....	.....G..	.C.....	.....C..C.
Bonobo	GC.....	T...A.....	.....	.C.....	.....C..C.
A-175	GC.....	T...A.....	.....	.C.....	.....C..C.
A-176	GC.....	T...A.....	.....	.C.....	.....C..C.
A-281	GC.....	T...A.....	.....	.C.....	.....C..C.
A-208	GC.....	T...A.....	.....	.C.....	.....C..C.
A-231	GC.....	T...A.....	.....	.C.....	.....C..C.
A-285	GC.....	T...A.....	.....	.C.....	.....C..C.
A-199	GC.....	T...A.....	.....	.C.....	.....C..C.
A-241	GC.....	T...A.....	.....	.C.....	.....C..C.
A-292	GC.....	T...A.....	.....	.C.....	.....C..C.
A-50	GC.....	T...A.....	.....	.C.....	.....C..C.
A-179	GC.....	T...A.....	.....	.C.....	.....C..C.
A-283	GC.....	T...A.....	.....	.C.....	.....C..C.
A-290	GC.....	T...A.....	.....	.C.....	.....C..C.
A-242	GC.....	T...A.....	.....	.C.....	.....C..C.
A-182	GC.....	T...A.....	.....	.C.....	.....C..C.
A-288	GC.....	T...A.....	.....	.C.....	.....C..C.
Pt-ref	GC.....	T...A.....	.....	.C.....	.....C..C.
A-291	GC.....	T...A.....	.....	.C.....	.....C..C.
A-282	GC.....	T...A.....	.....	.C.....	.....C..C.
A-287	GC.....	T...A.....	.....	.C.....	.....C..C.
A-235	GC.....	T...A.....	.....	.C.....	.....C..C.

	5170	5180	5190	5200	5210
Gorilla	GAAATAAACT	GACATGATTA	ACACCTCTAA	TCCCCTCCAC	CCTCCTCTCC
AFR1	....C..G..	A......C..	.....CT...	.T..A.....	....
AFR2	....C..G..	A......C..	.....CT...	.T..A.....	....
AFR3	....C..G..	A......C..	.....CT...	.T..A.....	....
AFR4	....C..G..	A......C..	.....CT...	.T..A.....	....
AFR5	....C..G..	A......C..	.....CT...	.T..A.....	....
EUR1	....C..G..	A......C..	.....CT...	.T..A.....	....
EUR2	....C..G..	A......C..	.....CT...	.T..A.....	....
EUR3	....C..G..	A......C..	.....CT...	.T..A.....	....
EUR4	....C..G..	A......C..	.....CT...	.T..A.....	....
EUR5	....C..G..	A......C..	.....CT...	.T..A.....	....
ASN1	....C..G..	A......C..	.....CT...	.T..A.....	....
ASN2	....C..G..	A......C..	.....CT...	.T..A.....	....
ASN3	....C..G..	A......C..	.....CT...	.T..A.....	....
ASN4	....C..GA.	A......C..	.....CT...	.T..A.....	....
ASN5	....C..GA.	A......C..	.....CT...	.T..A.....	....
AUS1	....C..G..	A......C..	.....CT...	.T..A.....	....
AUS2	....C..G..	A......C..	.....CT...	.T..A.....	....
AUS3	....C..G..	A......C..	.....CT...	.T..A.....	....
AUS4	....C..G..	A......C..	.....CT...	.T..A.....	....
AUS5	....C..G..	A......C..	.....CT...	.T..A.....	....
Hs-ref	....C..G..	A......C..	.....CT...	.T..A.....	....
Bonobo	....C..G..	A......C..	..T.......	....A.....	....
A-175	....C..G..	A.........	..T.......	....A.....	....
A-176	....C..G..	A.........	..T.......	....A.....	....
A-281	....C.....	A.........	..T.......	.T..A.....	....
A-208	....C..G..	A.........	..T.......	.T..A.....	....
A-231	....C..G..	A.........	..T.......	.T..A.....	....
A-285	....C..G..	A.........	..T.......	.T..A.....	....
A-199	....C..G..	A.........	..T.......	.T..A.....	....
A-241	....C..G..	A.........	..T.......	.T..A.....	....
A-292	....C..G..	A.........	..T.......	.T..A.....	....
A-50	....C..G..	A.........	..T.......	.T..A.....	....
A-179	....C..G..	A.........	..T.......	.T..A.....	....
A-283	....C..G..	A.........	..T.......	.T..A.....	....
A-290	....C..G..	A.........	..T.......	.T..A.....	....
A-242	....C..G..	A.........	..T.......	.T..A.....	....
A-182	....C..G..	A.........	..T.......	.T..A.....	....
A-288	....C..G..	A.........	..T..C....	.T..A.....	....
Pt-ref	....C..G..	A.........	..T..C....	.T..A.....	....
A-291	....C..G..	A.........	..T..C....	.T..A.....	....
A-282	....C..G..	A.........	..T..C....	.T..A.....	....
A-287	....C..G..	A.........	..T..C....	.T..A.....	....
A-235	....C.....	A.........	..T..C....	.T..A.....	....

```
                5220        5230        5240        5250        5260
                |           |           |           |           |
Gorilla         CTAGGAGGTC  TACCCCCACT  AACCGGCTTC  CTACCCAAAT  GGCTTATTAT
AFR1            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AFR2            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AFR3            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AFR4            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AFR5            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
EUR1            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
EUR2            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
EUR3            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
EUR4            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
EUR5            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
ASN1            ........C.  .G.....G..  .........T  T.G.......  ..G.C.....
ASN2            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
ASN3            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
ASN4            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
ASN5            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AUS1            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AUS2            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AUS3            ........C.  .G.....G..  .........T  T.G.......  ..ACC.....
AUS4            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
AUS5            ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
Hs-ref          ........C.  .G.....G..  .........T  T.G.......  ..GCC.....
Bonobo          ........C.  ..........  ..........  T.G.......  .AG....C..
A-175           ........C.  .G........  ..........  T.G.......  .AGC...C..
A-176           ........C.  .G........  ..........  T.G.......  .AGC...C..
A-281           ........C.  ..........  ..........  T.........  .AG....C..
A-208           ........C.  ..........  ..........  T.........  .AG....C..
A-231           ........C.  ..........  ..........  T.........  .AG....C..
A-285           ........C.  ..........  ..........  T.........  .AG....C..
A-199           ........C.  ..........  ..........  T.........  .AG....C..
A-241           ........C.  ..........  ..........  T.........  .AG....C..
A-292           ........C.  ..........  ..........  T.........  .AG....C..
A-50            ........C.  ..........  ..........  T.........  .AG....C..
A-179           ........C.  ..........  ..........  T.........  .AG....C..
A-283           ........C.  ..........  ..........  T.........  .AG....C..
A-290           ........C.  ..........  ..........  T.........  .AG....C..
A-242           ........C.  ..........  ..........  T.........  .AG....C..
A-182           ........C.  ..........  ...T......  T.........  .AG....C..
A-288           ........C.  ..........  ...T......  T.........  .AG....C..
Pt-ref          ........C.  ..........  ...T......  T.........  .AG....C..
A-291           ........C.  ..........  ...T......  T.........  .AG....C..
A-282           ........C.  .......G..  ...T......  T.........  .AG....C..
A-287           ........C.  .......G..  ...T......  T.........  .AG....C..
A-235           ........C.  .......G..  ...T......  T.........  .AG....C..
```

	5270	5280	5290	5300	5310
Gorilla	CGAAGAATTC	ACAAAAAATA	ACGACCTCAT	TACCCCCACC	ATTATGGCCA
AFR1	....	........C.	.TAG......	C.T.......	..C..A....
AFR2	....	........C.	.TAG......	C.T.......	..C..A....
AFR3	....	........C.	.TAG......	C.T.......	..C..A....
AFR4	....	........C.	.TAG......	C.T.......	..C..A....
AFR5	....	........C.	.TAG......	C.T.......	..C..A....
EUR1	....	........C.	.TAG......	C.T.......	..C..A....
EUR2	....	........C.	.TAG......	C.T.......	..C..A....
EUR3	....	........C.	.TAG......	C.T.......	..C..A....
EUR4	....	........C.	.TAG......	C.T.......	..C..A....
EUR5	....	........C.	.TAG......	C.T.......	..C..A....
ASN1	....	........C.	.TAG......	C.T.......	..C..A....
ASN2	....	........C.	.TAG......	C.T.......	..C..A....
ASN3	....	........C.	.TAG......	C.T.......	..C..A....
ASN4	....	........C.	.TAG......	C.T.......	..C..A....
ASN5	....	........C.	.TAG......	C.T.......	..C..A....
AUS1	....	........C.	.TAG......	C.T.......	..C..A....
AUS2	....	........C.	.TAG......	C.........	..C..A....
AUS3	....	........C.	.TAG......	C.T.......	..C..A....
AUS4	....	........C.	.TAG......	C.T.......	..C..A....
AUS5	....	........C.	.TAG......	C.T.......	..C..A....
Hs-ref	....	........C.	.TAG......	C.T.......	..C..A....
Bonobo	....	....	.TAG......	..T.......	.CC..A....
A-175	....	....	..AG......	C.T.......	..C..A....
A-176	....	....	.TAG......	C.T.......	..C..A....
A-281	....	....	.TAG......	C.T.......	..C..A....
A-208	....	....	.TAG......	C.TT......	..C..A....
A-231	....	....	.TA.......	C.TT......	..C..A....
A-285	....	....	.TAG......	C.TT......	..C..A....
A-199	....	....	.TAG......	C.TT......	..C..A....
A-241	....	....	.TAG......	C.TT......	..C..A....
A-292	....	....	.TAG......	C.TT......	..C..A....
A-50	....	....	.TAG......	C.TT......	..C..A....
A-179	....	....	.TAG......	C.TT......	..C..A....
A-283	....	....	.TAG......	C.TT......	..C..A....
A-290	....	....	.TAG......	C.TT......	..C..A....
A-242	....	....	.TAG......	C.TT......	..C..A....
A-182	....	....	.TAG......	C.T.......	..C..A....
A-288	....	....	.TAG......	C.T.......	..C..A....
Pt-ref	....	....	.TAG......	C.T.......	..C..A....
A-291	....	....	.TAG......	C.T.......	..C..A....
A-282	....	....	.TAG......	C.T.......	..C..A....
A-287	....	....	.TAG......	C.T.......	..C..A....
A-235	....	....	.TAG......	C.T.......	..C..A....

```
                5320        5330        5340        5350        5360
                |           |           |           |           |
Gorilla         TCATCACCCT  CCTCAACCTC  TATTTCTATC  TACGCCTAAT  TTACTCCACC
AFR1            C.........  ...T......  ..C.....C.  ..........  C.........
AFR2            C.........  ...T......  ..C.....C.  ..........  C.........
AFR3            C.........  ...T......  ..C.....C.  ..........  C.........
AFR4            C.........  ...T......  ..C.....C.  ..........  C.........
AFR5            C.........  ...T......  ..C.....C.  ..........  C.........
EUR1            C.........  ...T......  ..C.....C.  ..........  C.........
EUR2            C.........  ...T......  ..C.....C.  ..........  C.........
EUR3            C.........  ...T......  ..C.....C.  ..........  C.........
EUR4            C.........  ...T......  ..C.....C.  ..........  C.........
EUR5            C.........  ...T......  ..C.....C.  ..........  C.........
ASN1            C.........  ...T......  ..C.....C.  ..........  C.........
ASN2            C.........  ...T......  ..C.....C.  ..........  C.........
ASN3            C.........  ...T......  ..C.....C.  .G........  C.........
ASN4            C.........  ...T......  ..C.....C.  ..........  C.........
ASN5            C.........  ...T......  ..C.....C.  ..........  C.........
AUS1            C.........  ...T......  ..C.....C.  ..........  C.........
AUS2            C.........  ...T......  ..C.....C.  ..........  C.........
AUS3            C.........  ...T......  ..C.....C.  ..........  C.........
AUS4            C.........  ...T......  ..C.....C.  ..........  C.........
AUS5            C.........  ...T......  ..C.....C.  ..........  C.........
Hs-ref          C.........  ...T......  ..C.....C.  ..........  C.........
Bonobo          .......T..  ...T......  ........C.  .C........  C.........
A-175           .......T..  ..........  ........C.  ..........  C.........
A-176           .......T..  ..........  ........C.  ..........  C.........
A-281           .......T..  ...T......  ..C.....C.  ..........  C.........
A-208           .......T..  ...T......  ........C.  ..........  C.........
A-231           .......T..  ...T......  ........C.  ..........  C.........
A-285           .......T..  ...T......  ........C.  ..........  C.........
A-199           .......T..  ...T......  ........C.  ..........  C.........
A-241           .......T..  ...T......  ........C.  ..........  C.........
A-292           .......T..  ...T......  ........C.  ..........  C.........
A-50            .......T..  ...T......  ........C.  ..........  C........T
A-179           .......T..  ...T......  ........C.  ..........  C.........
A-283           .......T..  ...T......  ........C.  ..........  C.........
A-290           .......T..  ...T......  ........C.  ..........  C.........
A-242           .......T..  ...T......  ........C.  ..........  C.........
A-182           .......T..  ...T......  ........C.  ..........  C.........
A-288           .......T..  ...T......  ........C.  ..........  C.........
Pt-ref          .......T..  ...T......  ........C.  ..........  C.........
A-291           .......T..  ...T......  ........C.  ..........  C.........
A-282           .......T..  ...T......  ........C.  ..........  C.........
A-287           .......T..  ...T......  ........C.  ..........  C.........
A-235           .......T..  ...T......  ........C.  ..........  C.........
```

```
                5370        5380        5390        5400        5410
                |           |           |           |           |
Gorilla         TCTATCACAC  TACTACCCAT  ATCCAACAAC  GTAAAGATAA  AATGACAGCT
AFR1            ..A.......  ....C.....  ...T......  .....A....  .........T.
AFR2            ..A.......  ....C.....  ...T......  .....A....  .........T.
AFR3            ..A.......  ....C.....  ...T......  .....A....  .........T.
AFR4            ..A.......  ....C.....  ...T......  .....A....  .........T.
AFR5            ..A.......  ....C.....  ...T......  .....A....  .........T.
EUR1            ..A.......  ....C.....  ...T......  .....A....  .........T.
EUR2            ..A.......  ....C.....  ...T......  .....A....  .........T.
EUR3            ..A.......  ....C.....  ...T......  .....A....  .........T.
EUR4            ..A.......  ....C.....  ...T......  .....A....  .........T.
EUR5            ..A.......  ....C.....  ...T......  .....A....  .........T.
ASN1            ..A.......  ....C.....  ...T......  .....A....  .........T.
ASN2            ..A.......  ....C.....  ...T......  .....A....  .........T.
ASN3            ..A.......  ....C.....  ...T......  .....A....  .........T.
ASN4            ..A.......  ....C.....  ...T......  .....A....  .........T.
ASN5            ..A.......  ....C.....  ...T......  .....A....  .........T.
AUS1            ..A.......  ....C.....  ...T......  .....A....  .........T.
AUS2            ..A.......  ....C.....  ...T......  .....A....  .........T.
AUS3            ..A.......  ....C.....  ...T......  .....A....  .........T.
AUS4            ..A.......  ....C.....  ...T......  .....A....  .........T.
AUS5            ..A.......  ....C.....  ...T......  .....A....  .........T.
Hs-ref          ..A.......  ....C.....  ...T......  .....A....  .........T.
Bonobo          ..G.......  ....T..T..  ......T..T  .....A....  ........AT.
A-175           ..A..T....  ....T..T..  ......T...  .....A....  ........AT.
A-176           ..A..T....  ....T..T..  ......T...  .....A....  ........AT.
A-281           ..A..T....  ....T..T..  ......T...  .....A....  ........AT.
A-208           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-231           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-285           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-199           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-241           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-292           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-50            ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-179           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-283           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-290           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-242           ..A..T....  ....T..T..  ...T..T...  .....A....  ........AT.
A-182           ..A..T....  ....T.....  ...T..T...  .....A....  ........AT.
A-288           ..A..T....  ....T.....  ...T..T...  .....A....  ........AT.
Pt-ref          ..A..T....  ....T.....  ...T..T...  .....A....  ........AT.
A-291           ..A..T....  ....T.....  ...T..T...  .....A....  ........AT.
A-282           ..A..T....  ....T.....  ...T..T...  .....A....  ........AT.
A-287           ..A..T....  ....T.....  ...T..T...  .....A....  ........AT.
A-235           ..A..T....  ....T.....  ...T..T...  .....A....  ........AT.
```

```
                5420        5430        5440        5450        5460
                |           |           |           |           |
Gorilla         CGAATATACT  AAACCCACTC  CCTTCCTCCC  TACACTTATC  ACACTCACCA
AFR1            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AFR2            T...C....A  ........C.  .AC.......  C.....C...  G.C..T....
AFR3            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AFR4            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AFR5            T...C....A  ........C.  .AC.......  C.....C...  G.C..T....
EUR1            T...C....A  ........C.  .A........  C.....C...  G.C..T....
EUR2            T...C....A  ........C.  .A........  C.....C...  G.C..T....
EUR3            T...C....A  ........C.  .A........  C.....C...  G.C..T....
EUR4            T...C....A  ........C.  .A........  C.....C...  G.C..T....
EUR5            T...C....A  ........C.  .A........  C.....C...  G.C..T....
ASN1            T...C....A  ........C.  .A........  C.....C...  G.C..T....
ASN2            T...C....A  ........C.  .AC.......  C.....C...  G.C..T....
ASN3            T...C....A  ........C.  .A........  C.....C...  ..C..T....
ASN4            T...C....A  ........C.  .A........  C.....C...  G.C..TG...
ASN5            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AUS1            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AUS2            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AUS3            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AUS4            T...C....A  ........C.  .A........  C.....C...  G.C..T....
AUS5            T...C....A  ........C.  .A........  C.....C...  G.C..T....
Hs-ref          T...C....A  ........C.  .A........  C.....C...  G.C..T....
Bonobo          T...C....A  .....T..C.  ..........  ......C...  ..C..T....
A-175           ....C....A  ........C.  T.........  ......C...  ..C..T....
A-176           ....C....A  ........C.  T.........  ......C...  ..C..T....
A-281           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-208           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-231           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-285           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-199           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-241           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-292           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-50            ....C....A  ........C.  ..........  ......C...  ..C..T....
A-179           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-283           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-290           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-242           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-182           ....C....A  ........C.  .......T..  ......C...  ..C..T....
A-288           ....C....A  ........C.  ..........  ......C...  ..C..T....
Pt-ref          ....C....A  ........C.  ..........  ......C...  ..C..T....
A-291           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-282           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-287           ....C....A  ........C.  ..........  ......C...  ..C..T....
A-235           ....C....A  ........C.  ..........  ......C...  ..C..T....
```

```
                5470        5480        5490        5500        5510
                |           |           |           |           |
Gorilla         CCCTACTTCT ACCCATCTCC CCCTTCATAC TAATAGTTCT A
AFR1            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AFR2            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AFR3            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AFR4            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AFR5            .G.....C.. ...T..... ..T..T.... .....A.CT. .
EUR1            .G.....C.. ...T..... ..T..T.... .....A.CT. .
EUR2            .G.....C.. ...T..... ..T..T.... .....A.CT. .
EUR3            .G.....C.. ...T..... ..T..T.... .....A.CT. .
EUR4            .G.....C.. ...T..... ..T..T.... .....A.CT. .
EUR5            .G.....C.. ...T..... ..T..T.... .....A.CT. .
ASN1            .G.....C.. ...T..... ..T..T.... .....A.CT. .
ASN2            .G.....C.. ...T..... ..T..T.... .....A.CT. .
ASN3            .G.....C.. ...T..... ..T..T.... .....A.CT. .
ASN4            .G.....C.. ...T..... ..T..T.... .....A.CT. .
ASN5            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AUS1            .G.....C.. ...T..... .....T.... .....A.CT. .
AUS2            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AUS3            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AUS4            .G.....C.. ...T..... ..T..T.... .....A.CT. .
AUS5            .G.....C.. ...T..... ..T..T.... .....A.CT. .
Hs-ref          .G.....C.. ...T..... ..T..T.... .....A.CT. .
Bonobo          .A..... ..... ..T..... ....GA.CT. .
A-175           .A..... ..... ..T..... .....A.CT. .
A-176           .A..G.... ..... ..T..... .....A.CT. .
A-281           .A..G.... ..... ..T..... .....A.CT. .
A-208           .A..G.... ..... ..T..... .....A.CT. .
A-231           .A..G.... ..... ..T..... .....A.CT. .
A-285           .A..G.... ..... ..T..... .....A.CT. .
A-199           .A..G.... ..... ..T..... .....A.CT. .
A-241           .A..G.... ..... ..T..... .....A.CT. .
A-292           .A..G.... ..... ..T..... .....A.CT. .
A-50            .A..G.... ..... ..T..... .....A.CT. .
A-179           .A..G.... ..... ..T..... .....A.CT. .
A-283           .A..G.... ..... ..T..... .....A.CT. .
A-290           .A..G.... ..... ..T..... .....A.CT. .
A-242           .A..G.... ..... ..... .....A.CT. .
A-182           .A..G.... ..... ..... .....A.CT. .
A-288           .A..G.... ..... ..... .....A.CT. .
Pt-ref          .A..G.... ..... ..... .....A.CT. .
A-291           .A..G.... ..... ..... .....A.CT. .
A-282           .A..G.... ..... ..... .....A.CT. .
A-287           .A..G.... ..... ..... .....A.CT. .
A-235           .A..G.... ..... ..... .....A.CT. .
```

APPENDIX C

Nucleotide sequence alignment of the mitochondrial cytochrome *b* (*cyt b*) gene from humans, chimpanzees, bonobo and gorilla. Nucleotide positions are numbered according to Anderson *et al.* (1981). A dot (.) indicates sequence identity and a ? denotes missing data. Human sequences are from Africa (AFR), Europe (EUR), Asia (ASN) and Australia (AUS). Hs-ref is a previously published human sequence (Anderson *et al.* 1981). The majority of the chimpanzee sequences are from the *Pan troglodytes verus* subspecies, with the exception of A-175 and A-176 which are from *P. t. troglodytes* and/or *P. t. schweinfurthii* (see Section 3.4.1; Wise *et al.* 1997). Pt-ref is a previously published chimpanzee sequence (Horai *et al.* 1995). The bonobo and gorilla sequences are also published in Horai *et al.* (1995).

```
                14747       14757       14767       14777       14787
                |           |           |           |           |
Gorilla         ATGACCCCTA  TACGCAAAAC  TAACCCACTA  GCAAAACTAA  TTAACCACTC
AFR1            ........A.  .........T  ......C...  AT....T...  ..........
AFR2            ........A.  .........T  ......C...  AT....T...  ..........
AFR6            ........A.  .........T  ......C...  AT....T...  ..........
AFR7            ........A.  .........T  ......C...  AT....T...  ..........
AFR8            ........A.  .........T  ......C...  AT....T...  ..........
EUR1            ........A.  .........T  ......C...  AT....T...  ..........
EUR2            ........A.  .........T  ......C...  AT....T...  ..........
EUR3            ........A.  .........T  ......C...  AT....T...  ..........
EUR4            ........A.  ..........  ......C...  AT....T...  ..........
EUR5            ........A.  .........T  ......C...  AT....T...  ..........
ASN1            ........A.  .........T  ......C...  AT....T...  ..........
ASN2            ........A.  .........T  ......C...  AT........  ..........
ASN3            ........A.  .........T  ......C...  AT........  ..........
ASN6            ........A.  .........T  ......C...  AT....T...  ..........
ASN7            ........A.  .........T  ......C...  AT........  ..........
AUS1            ........A.  .........T  ......C...  AT........  ..........
AUS3            ........A.  ..?......T  ......C...  AT....T...  ..........
AUS4            ........A.  .........T  ......C...  AT....T...  ..........
AUS5            ........A.  .........T  ......C...  AT........  ..........
AUS6            ........A.  .........T  ......C...  AT........  ..........
Hs-ref          ........A.  .........T  ......C...  AT....T...  ..........
Bonobo          ........A.  C........T  C.........  AT....T...  ....T.....
A-175           ........A.  C........T  ..........  AT....T...  ....T.....
A-176           ........A.  C..........  ...T......  AT....T...  ....T.....
A-281           ........A.  C........T  ..........  AT....T...  ....T.....
A-208           ........A.  C........T  ..........  AT....T...  ....T.....
A-231           ........A.  C........T  ..........  AT....T...  ....T.....
A-285           ........A.  C........T  ..........  AT....T...  ....T.....
A-199           ........A.  C........T  ..........  AT....T...  ....T.....
A-241           ........A.  C........T  ..........  AT....T...  ....T.....
A-292           ........A.  C........T  ..........  AT....T...  ....T.....
A-50            ........A.  C........T  ..........  AT....T...  ....T.....
A-179           ........A.  C........T  ..........  AT....T...  ....T.....
A-283           ........A.  C........T  ..........  AT....T...  ....T.....
A-290           ........A.  C........T  ..........  AT....T...  ....T.....
A-242           ........A.  C........T  ..........  AT....T...  ....T.....
A-182           ........A.  C........T  ..........  AT....T...  ....T.....
A-288           ........G.  C........T  ..........  AT....T...  ....T.....
Pt-ref          ........G.  C........T  ..........  AT....T...  ....T.....
A-291           ........G.  C........T  ..........  AT....T...  ....T.....
A-282           ........G.  C........T  ..........  AT....T...  ....T.....
A-287           ........A.  C........T  ..........  AT....T...  ....T.....
A-235           ........A.  C........T  ..........  AT....T...  ....T.....
```

	14797	14807	14817	14827	14837
	|	|	|	|	|
Gorilla	ATTCATTGAC	CTCCCTACCC	CGTCCAACAT	CTCCACATGA	TGAAACTTCG
AFR1	......C...	.....C....	.A........	....G.....	....
AFR2	......C...	.....C....	.A........	....G.....	....
AFR6	......C...	.....C....	.A........	....G.....	....
AFR7	......C...	.....C....	.A........	....G.....	....
AFR8	......C...	.....C....	.A........	....G.....	....
EUR1	.C....C...	.....C....	.A........	....G.....	....
EUR2	......C...	.....C....	.A........	....G.....	....
EUR3	......C...	.....C....	.A........	....G.....	....
EUR4	......C...	.....C....	.A........	....G.....	....
EUR5	......C...	.....C....	.A........	....G.....	....
ASN1	......C...	.....C....	.A........	....G.....	....
ASN2	......C...	.....C....	.A........	....G.....	....
ASN3	......C...	.....C....	.A........	....G.....	....
ASN6	......C...	.....C....	.A........	....G.....	....
ASN7	......C...	.....C....	.A........	....G.....	....
AUS1	......C...	.....C....	.A........	....G.....	....
AUS3	......C...	.....C....	.A........	....G.....	....
AUS4	......C...	.....C....	.A........	....G.....	....
AUS5	......C...	.....C....	.A........	....G.....	....
AUS6	......C...	.....C....	.A........	....G.....	....
Hs-ref	......C...	.....C....	.A........	....G.....	....
Bonobo	...T..C...	.....C....	.A.....T..	T.........	....
A-175	...T..C...	.....C....	.A........	T...G.....	..G.......
A-176	...T..C...	.....C....	.A........	T...G.....	..G.......
A-281	...T..C...	.....C....	.A........	T...G.....	..G.......
A-208	...T..C...	.....C....	.A........	T...G.....	..G.......
A-231	...T..C...	.....C....	.A........	T...G.....	..G.......
A-285	...T..C...	.....C....	.A........	T...G.....	..G.......
A-199	...T..C...	.....C....	.A........	T...G.....	..G.......
A-241	...T..C...	.....C....	.A........	T...G.....	..G.......
A-292	...T..C...	.....C....	.A........	T...G.....	..G.......
A-50	...T..C...	.....C....	.A........	T...G.....	..G.......
A-179	...T..C...	.....C....	.A........	T...G.....	..G.......
A-283	...T..C...	.....C....	.A........	T...G.....	..G.......
A-290	...T..C...	.....C....	.A........	T...G.....	..G.......
A-242	...T..C...	.....C....	.A........	T...G.....	..G.......
A-182	...T......	.....C....	.A........	T...G.....	..G.......
A-288	...T..C...	.....C....	.A........	T...G.....	..G.......
Pt-ref	...T..C...	.....C....	.A........	T...G.....	..G.......
A-291	...T..C...	.....C....	.A........	T...G.....	..G.......
A-282	...T..C...	.....C....	.A........	T...G.....	..G.......
A-287	...T..C...	.....C....	.A........	T...G.....	..G.......
A-235	...T..C...	.....C....	.A........	T...G.....	..G.......

	14847	14857	14867	14877	14887
Gorilla	GCTCACTCCT	TGGTGCCTGC	TTAATCCTTC	AAATCACCAC	AGGGCTATTC
AFR1	....	...C......	C.G.....C.	....	...A......
AFR2	....	...C......	C.G.....C.	....	...A......
AFR6	....	...C......	C.G.....C.	....	...A......
AFR7	....	...C......	C.G.....C.	....	...A......
AFR8	....	...C......	C.G.....C.	....	...A......
EUR1	....	...C......	C.G.....C.	....	...A......
EUR2	....	...C......	C.G.....C.	....	...A......
EUR3	....	...C......	C.G.....C.	....	...A......
EUR4	....	...C......	C.G.....C.	....	...A......
EUR5	....	...C......	C.G.....C.	....	...A......
ASN1	....	...C......	C.G.....C.	....	...A......
ASN2	....	...C......	C.G.....C.	....	...A......
ASN3	....	...C......	C.G.....C.	....	...A......
ASN6	....	...C......	C.G.....C.	....	...A......
ASN7	....	...C......	C.G.....C.	....	...A......
AUS1	....	...C......	C.G.....C.	....	...A......
AUS3	....	...C......	C.G.....C.	....	...A......
AUS4	....	...C......	C.G.....C.	....	...A......
AUS5	....	...C......	C.G.....C.	....	...A......
AUS6	....	...C......	C.G.....C.	....	...A......
Hs-ref	....	...C......	C.G.....C.	....	...A......
Bonobo	.......T..	C..C......	C.........	....	...A......
A-175	....?..T..	C..C......	C.........	....	...AT.....
A-176	.......T..	C..C......	C.........	....	...AT.....
A-281	.......T..	C..C......	C.........	....T.....	...AT.....
A-208	.......T..	C..C......	C.........	....T.....	...AT.....
A-231	.......T..	C..C......	C.........	....T.....	...AT.....
A-285	.......T..	C..C......	C.........	....T.....	...AT.....
A-199	.......T..	C..C......	C.........	....T.....	...AT.....
A-241	.......T..	C..C......	C.........	....T.....	...AT.....
A-292	.......T..	C..C......	C.........	....T.....	...AT.....
A-50	.......T..	C..C......	C.........	....T.....	...AT.....
A-179	.......T..	C..C......	C.........	....T.....	...AT.....
A-283	.......T..	C..C......	C.........	....T.....	...AT.....
A-290	.......T..	C..C......	C.........	....T.....	...AT.....
A-242	.......T..	C..C......	C.........	....T.....	...AT.....
A-182	.......T..	C..CA.....	C.........	....T.....	...AT.....
A-288	.......T..	C..C......	C.........	....T.....	...AT.....
Pt-ref	.......T..	C..C......	C.........	....T.....	...AT.....
A-291	.......T..	C..C......	C.........	....T.....	...AT.....
A-282	.......T..	C..C......	C.........	....T.....	...AT.....
A-287	.......T..	C..C......	C.........	....T.....	...AT.....
A-235	.......T..	C..C......	C.........	....T.....	...AT.....

```
                14897      14907      14917      14927      14937
                |          |          |          |          |
Gorilla         CTAGCCATAC ACTACTCACC TGATGCCTCA ACCGCCTTCT CATCAATTGC
AFR1            ........G. ....T..... A..C...... .........T. .......C..
AFR2            ........G. .......... A..C...... .........T. .......C..
AFR6            ........G. ....T..... A..C...... .........T. .......C..
AFR7            .......... .......... A..C...... .........T. .......C..
AFR8            ........G. .......... A..C...... .........T. .......C..
EUR1            ........G. .......... A..C...... .........T. .......C..
EUR2            .......... .......... A..C...... .........T. .......C..
EUR3            ........G. .......... A..C...... .........T. .......C..
EUR4            ........G. .......... A..C...... .........T. .......C..
EUR5            .......... .......... A..C...... .........T. .......C..
ASN1            ........G. .......... A..C...... .........T. .......C..
ASN2            ........G. .......... A..C...... .........T. .......C..
ASN3            ........G. .......... A..C...... .........T. .......C..
ASN6            ........G. .......... A..C...... .........T. .......C..
ASN7            ........G. .......... A..C...... .........T. .......C..
AUS1            ........G. .......... A..C...... .........T. .......C..
AUS3            ........G. .......... A..C...... .........T. .......C..
AUS4            ........G. .......... A..C...... .........T. .......C..
AUS5            ........G. .......... A..C...... .........T. .......C..
AUS6            ........G. .......... A..C...... .........T. .......C..
Hs-ref          ........G. .......... A..C...... .........T. .......C..
Bonobo          .....T.... .......... A..C...... .......... ....G..C..
A-175           .......... .......... A..C...... .......... .G..G..C..
A-176           .......... .......... A..C...... .......... .G..G..C..
A-281           .....T.... .......... A..C...... .......... .G..G..C..
A-208           .....T.... .......... A..C...... .......... .G..G..C..
A-231           .....T.... .......... A..C...... .......... .G..G..C..
A-285           .....T.... .......... A..C...... .......... .G..G..C..
A-199           .....T.... .......... A..C...... .......... .G..G..C..
A-241           .....T.... .......... A..C...... .......... .G..G..C..
A-292           .....T.... .......... A..C...... .......... .G..G..C..
A-50            .....T.... .......... A..C...... .......... .G..G..C..
A-179           .....T.... .......... A..C...... .......... .G..G..C..
A-283           .....T.... .......... A..C...... .......... .G..G..C..
A-290           .....T.... .......... A..C...... .......... .G..G..C..
A-242           .....T.... .......... A..C...... .......... .G..G..C..
A-182           .....T.... .......... A..C...... .......... .G..G..C..
A-288           .....T.... .......... A..C...... .......... .G..G..C..
Pt-ref          .....T.... .......... A..C...... .......... .G..G..C..
A-291           .....T.... .......... A..C...... .......... .G..G..C..
A-282           .....T.... .......... A..C...... .......... .G..G..C..
A-287           .....T.... .......... A..C...... .......... .G..G..C..
A-235           .....T.... .......... A..C...... .......... .G..G..C..
```

	14947	14957	14967	14977	14987
	|	|	|	|	|
Gorilla	CCACATCACC	CGAGATGTAA	ACTATGGCTG	AACCATCCGC	TACCTCCACG
AFR1	.....T	.....C....	.T.....	..T.....	.....T....
AFR2	.....T	.....C....	.T.....	..T.....	.....T....
AFR6	.....T	.....C....	.T.....	..T.....	.....T....
AFR7	.....T	.....C....	.T.....	..T.....	.....T....
AFR8	.....T	.....C....	.T.....	..T.....	.....T....
EUR1	.....T	.....C....	.T.....	..T.....	.....T....
EUR2	.....T	.....C....	.T.....	..T.....	.....T....
EUR3	.....	.....C....	.T.....	..T.....	.....T....
EUR4	.....T	.....C....	.T.....	..T.....	.....T....
EUR5	.....T	.....C....	.T.....	..T.....	.....T....
ASN1	.....T	.....C....	.T.....	..T.....	.....T....
ASN2	.....T	.....C....	.T.....	..T.....	.....T....
ASN3	.....T	.....C....	.T.....	..T.....	.....T....
ASN6	.....T	.....C....	.T.....	..T.....	.....T....
ASN7	.....T	.....C....	.T.....	..T.....	.....T....
AUS1	.....T	.....C....	.T.....	..T.....	.....T....
AUS3	.....T	.....C....	.T.....	..T.....	.....T....
AUS4	.....T	.....C....	.T.....	..T.....	.....T....
AUS5	.....T	.....C....	.T.....	..T.....	.....T....
AUS6	.....T	.....C....	.T.....	..T.....	.....T....
Hs-ref	.....T	.....C....	.T.....	..T.....	.....T....
Bonobo	.....T...	.....C....	.....T..	..T.....	.....T....
A-175	.....	.....C....	.....	G.T.....	.....T....
A-176	.....	.....C....	.....T..	G.T.....	.....T....
A-281	.....	.....C....	.....	G.T.....	.....
A-208	.....	.....C....	.....T..	G.T.....	.....
A-231	.....	.....C....	.....T..	G.T.....	.....
A-285	.....	.....C....	.....T..	G.T.....	.....
A-199	.....	.....C....	.....T..	G.T.....	.....
A-241	.....	.....C....	.....T..	G.T.....	.....
A-292	.....	.....C....	.....T..	G.T.....	.....
A-50	.....	.....C....	.....T..	G.T.....	.....
A-179	.....	.....C....	.....T..	G.T.....	.....
A-283	.....	.....C....	.....T..	G.T.....	.....
A-290	.....	.....C....	.....T..	G.T.....	.....
A-242	.....	.....C....	.....T..	G.T.....	.....
A-182	.....	.....C....	.....T..	G.T.....	.....
A-288	.....	.....C....	.....T..	G.T.....	.....
Pt-ref	.....	.....C....	.....T..	G.T.....	.....
A-291	.....	.....C....	.....T..	G.T.....	.....
A-282	.....	.....C....	.....T..	G.T.....	.....
A-287	.....	.....C....	.....T..	G.T.....	.....
A-235	.....	.....C....	.....T..	G.T.....	.....

	14997	15007	15017	15027	15037
Gorilla	CTAACGGCGC	CTCAATATTC	TTCATTTGCC	TCTTTCTACA	CATCGGACGG
AFR1	.C..T.....	....	..T..C....	....C.....	......G..A
AFR2	.C..T.....	....	..T..C....	....C.....	......G..A
AFR6	.C..T.....	....	..T..C....	....C.....	......G..A
AFR7	.C..T.....	....	..T..C....	....C.....	......G..A
AFR8	.C..T.....	....	..T..C....	....C.....	......G..A
EUR1	.C..T.....	....	..T..C....	....C.....	......G..A
EUR2	.C..T.....	....	..T..C....	....C.....	......G..A
EUR3	.C..T.....	....	..T..C....	....C.....	......G..A
EUR4	.C..T.....	....	..T..C....	....C.....	......G..A
EUR5	.C..T.....	....	..T..C....	....C.....	......G..A
ASN1	.C..T.....	....	..T..C....	....C.....	......G..A
ASN2	.C..T.....	....	..T..C....	....C.....	.........A
ASN3	.C..T.....	....	..T..C....	....C.....	.........A
ASN6	.C..T.....	....	..T..C....	....C.....	......G..A
ASN7	.C..T.....	....	..T..C....	....C.....	.........A
AUS1	.C..T.....	....	..T..C....	....C.....	.........A
AUS3	.C..T.....	....	..T..C....	....C.....	......G..A
AUS4	.C..T.....	....	..T..C....	....C.....	......G..A
AUS5	.C..T.....	....	..T..C....	....C.....	.........A
AUS6	.C..T.....	....	..T..C....	....C.....	.........A
Hs-ref	.C..T.....	....	..T..C....	....C.....	......G..A
Bonobo	....	.......C.T	.....C....	....C.....	.G....T..A
A-175	.C..T.....	.........T	..T..C....	....C.....	......T..A
A-176	....T.....	.........T	..T..C....	....C.....	......T..A
A-281	....	.........T	..T..C....	....C.....	......T..A
A-208	....	.........T	..T..C....	....C.....	......T..A
A-231	....	.........T	..T..C....	....C.....	......T..A
A-285	....	.........T	..T..C....	....C.....	......T..A
A-199	....	.........T	..T..C....	....C.....	......T..A
A-241	....	.........T	..T..C....	....C.....	......T..A
A-292	....	.........T	..T..C....	....C.....	......T..A
A-50	....	.........T	..T..C....	....C.....	......T..A
A-179	....	.........T	..T..C....	....C.....	......T..A
A-283	....	.........T	..T..C....	....C.....	......T..A
A-290	....	.........T	..T..C....	....C.....	......T..A
A-242	....	.........T	..T..C....	....C.....	......T..A
A-182	....	.........T	..T..C....	....C.....	......C..A
A-288	....	.........T	..T..C....	....C.....	......C..A
Pt-ref	....	.........T	..T..C....	....C.....	......C..A
A-291	....	.........T	..T..C....	....C.....	......C..A
A-282	....	.........T	..T..C....	....C.....	......C..A
A-287	....	.........T	..T..C....	....C.....	......C..A
A-235	....	.........T	..T..C....	....C.....	......C..A

	15047	15057	15067	15077	15087
Gorilla	GGCCTATACT	ACGGCTCATT	TCTCCACCAA	GAAACCTGAA	ACATCGGCAT
AFR1	.....T.	....A.....	....T..TC.	.....	.....
AFR2	.....T.	....A.....	....T..TC.	.....	.....
AFR6	.....T.	....A.....	....T..TC.	.....	.....
AFR7	.....T.	....A.....	....T..TC.	.....	.....
AFR8	.....T.	....A.....	....T..TC.	.....	.....
EUR1	.....T.	....A.....	....T..TC.	.....	.....
EUR2	.....T.	....A.....	....T..TC.	.....	.....
EUR3	.....T.	....A.....	....T..TC.	.....	.....
EUR4	.....T.	....A.....	....T..TC.	.....	.....
EUR5	.....T.	....A.....	....T..TC.	.....	.....
ASN1	.....T.	....A.....	....T..TC.	.....	.....
ASN2	.....T.	....A.....	....T..TC.	.....	.....
ASN3	.....T.	....A.....	....T..TC.	.....	.....
ASN6	.....T.	....A.....	....T..TC.	.....	.....
ASN7	.....T.	....A.....	....T..TC.	.....	.....
AUS1	.....T.	....A.....	....T..TC.	.....	.....
AUS3	.....T.	....A.....	....T..TC.	.....	.....
AUS4	.....T.	....A.....	....T..TC.	.....	.....
AUS5	.....T.	....A.....	....T..TC.	.....	.....
AUS6	.....T.	....A.....	....T..TC.	.....	.....
Hs-ref	.....T.	....A.....	....T..TC.	.....	.....
Bonobo	.....T.	.....	....T...T.	.....	....T.....
A-175	.....T.	.....	....T...T.	.....	....T.....
A-176	.....T.	.....	....T...T.	.....	....T.....
A-281	..T.....T.	.....	...TT...T.	.....	....T.....
A-208	..T.....T.	.....	....T...T.	.....	....T.....
A-231	..T.....T.	.....	....T...T.	.....	....T.....
A-285	..T.....T.	.....	....T...T.	.....	....T.....
A-199	..T.....T.	.....	....T...T.	.....	....T.....
A-241	..T.....T.	.....	....T...T.	.....	....T.....
A-292	..T.....T.	.....	....T...T.	.....	....T.....
A-50	..T.....T.	.....	....T...T.	.....	....T.....
A-179	..T.....T.	.....	....T...T.	.....	....T.....
A-283	..T.....T.	.....	....T...T.	.....	....T.....
A-290	..T.....T.	.....	....T...T.	.....	....T.....
A-242	..T.....T.	.....	....T...T.	.....	....T.....
A-182	..T.....T.	.....	....T...T.	.....	....T.....
A-288	..T.....T.	.....	....T...T.	.....	....T.....
Pt-ref	..T.....T.	.....	....T...T.	.....	....T.....
A-291	..T.....T.	.....	....T...T.	.....	....T.....
A-282	..T.....T.	.....	....T...T.	.....	....T.....
A-287	..T.....T.	.....	....T...T.	.....	....T.....
A-235	..T.....T.	.....	....T...T.	.....	....T.....

```
                15097       15107       15117       15127       15137
                |           |           |           |           |
Gorilla         CATCCTCCTA CTCACAACCA TAGCAGCAGC CTTCATAGGC TATGTCCTCC
AFR1            T........G ..TG....T. .....A.... .......... ..........
AFR2            T........G ..TG....T. .....A.... .......... ..........
AFR6            T........G ..TG....T. .....A.... .......... ..........
AFR7            T........G ..TG....T. .....A.... .......... ..........
AFR8            T........G ..TG....T. .....A.... .......... ..........
EUR1            T........G ..TG....T. .....A.... .......... ..........
EUR2            T........G ..TG....T. .....A.... .......... ..........
EUR3            T........G ..TG....T. .....A.... .......... ..........
EUR4            T........G ..TG....T. .....A.... .......... ..........
EUR5            T........G ..TG....T. .....A.... .......... ..........
ASN1            T........G ..TG....T. .....A.... .......... ..........
ASN2            T........G ..TG....T. .....A.... .......... ..........
ASN3            T........G ..TG....T. .....A.... .......... ..........
ASN6            T........G ..TG....T. .....A.... .......... ..........
ASN7            T........G ..TG....T. .....A.... .......... ..........
AUS1            T........G ..TG....T. .....A.... .......... ..........
AUS3            T........G ..TG....T. .....A.... .......... ..........
AUS4            T........G ..TG....T. .....A.... .......... ..........
AUS5            T........G ..TG....T. .....A.... .......... ..........
AUS6            T........G ..TG....T. .....A.... .......... ..........
Hs-ref          T........G ..TG....T. .....A.... .......... ..........
Bonobo          .......T.G .......... .....A.... ...T..G... ..........
A-175           .......T.G .......... .....A.... ...T..G... ..........
A-176           .......T.G .......... .....A.... ...T..G... ..........
A-281           T......T.G .......... .....A.... ...T..G... ..........
A-208           T......T.G .......... .....A.... ...T..G... ..........
A-231           T......T.G .......... .....A.... ...T..G... ..........
A-285           T......T.G .......... .....A.... ...T..G... ..........
A-199           T......T.G .......... .....A.... ...T..G... ..........
A-241           T......T.G .......... .....A.... ...T..G... ..........
A-292           T......T.G .......... .....A.... ...T..G... ..........
A-50            T......T.G .......... .....A.... ...T..G... ..........
A-179           T......T.G .......... .....A.... ...T..G... ..........
A-283           T......T.G .......... .....A.... ...T..G... ..........
A-290           T......T.G .......... .....A.... ...T..G... ..........
A-242           T......T.G .......... .....A.... ...T..G... ..........
A-182           T......T.G .......... .....A.... ...T..G... ..........
A-288           T......T.G .......... .....A.... ...T..G... ..........
Pt-ref          T......T.G .......... .....A.... ...T..G... ..........
A-291           T......T.G .......... .....A.... ...T..G... ..........
A-282           T......T.G .......... .....A.... ...T..G... ..........
A-287           T......T.G .......... .....A.... ...T..G... ..........
A-235           T......T.G .......... .....A.... ...T..G... ..........
```

```
                15147       15157       15167       15177       15187
                |           |           |           |           |
Gorilla         CATGAGGCCA  AATATCCTTC  TGAGGAGCCA  CAGTAATCAC  AAACTTGCTA
AFR1            .G........  ......A...  .....G....  .......T..  ......A...
AFR2            .G........  ......A...  .....G....  .......T..  ......A...
AFR6            .G........  ......A...  .....G....  .......T..  ......A...
AFR7            .G........  ......A...  .....G....  .......T..  ......A...
AFR8            .G........  ......A...  .....G....  .......T..  ......A...
EUR1            .G........  ......A...  .....G....  .......T..  ......A...
EUR2            .G........  ......A...  .....G....  .......T..  ......A...
EUR3            .G........  ......A...  .....G....  .......T..  ......A...
EUR4            .G........  ......A...  .....G....  .......T..  ......A...
EUR5            .G........  ......A...  .....G....  .......T..  ......A...
ASN1            .G........  G.....A...  .....G....  .......T..  ......A...
ASN2            .G........  ......A...  .....G....  .......T..  ......A...
ASN3            .G........  ......A...  .....G....  .......T..  ......A...
ASN6            .G........  ......A...  .....G....  .......T..  ......A...
ASN7            .G........  ......A...  .....G....  .......T..  ......A...
AUS1            .G........  ......A...  .....G....  .......T..  ......A...
AUS3            .G........  ......A...  .....G....  .......T..  ......A...
AUS4            .G........  ......A...  ..........  .......T..  ......A...
AUS5            .G........  ......A...  .....G....  .......T..  ......A...
AUS6            .G........  ......A...  .....G....  .......T..  ......A...
Hs-ref          .G........  ......A...  .....G....  .......T..  ......A...
Bonobo          ..........  ..........  .....G....  .......T..  ....C.A..G
A-175           ..........  ..........  ..........  .......T..  ....C.A...
A-176           ..........  ..........  ..........  .......T..  ....C.A...
A-281           ..........  ..........  ..........  .......T..  ....C.A..G
A-208           ..........  ..........  ..........  .......T..  ....C.A..G
A-231           ..........  ..........  ..........  .......T..  ....C.A..G
A-285           ..........  ..........  ..........  .......T..  ....C.A..G
A-199           ..........  ..........  ..........  .......T..  ....C.A..G
A-241           ..........  ..........  ..........  .......T..  ....C.A..G
A-292           ..........  ..........  ..........  .......T..  ....C.A..G
A-50            ..........  ..........  ..........  .......T..  ....C.A..G
A-179           ..........  ..........  ..........  .......T..  ....C.A..G
A-283           ..........  ..........  ..........  .......T..  ....C.A..G
A-290           ..........  ..........  ..........  .......T..  ....C.A..G
A-242           ..........  ..........  ..........  .......T..  ....C.A..G
A-182           ..........  ..........  ..........  .......T..  ....C.A..G
A-288           ..........  ..........  ..........  .......T..  ....C.A..G
Pt-ref          ..........  ..........  ..........  .......T..  ....C.A..G
A-291           ..........  ..........  ..........  .......T..  ....C.A..G
A-282           ..........  ..........  ..........  .......T..  ....C.A..G
A-287           ..........  ..........  ..........  .......T..  ....C.A..G
A-235           ..........  ..........  ..........  .......T..  ....C.A..G
```

	15197	15207	15217	15227	15237
Gorilla	TCCGCCATCC	CGTACATCGG	AACAGACCTA	GTCCAATGAG	TTTGAGGTGG
AFR1	....	.A.....T..	G.........	..T......A	.C.....A..
AFR2	....	.A.....T..	G.........	..T......A	.C.....A..
AFR6	....	.A.....T..	G........G	..T......A	.C.....A..
AFR7	....	.A.....T..	G.........	..T......A	.C.....A..
AFR8	....	.A.....T..	G.........	..T......A	.C.....A..
EUR1	....	.A.....T..	G.........	..T......A	.C.....A..
EUR2	....	.A.....T..	G.........	..T......A	.C.....A..
EUR3	....	.A.....T..	G.........	..T......A	.C.....A..
EUR4	....	.A.....T..	G.........	..T......A	.C.....A..
EUR5	....	.A.....T..	G.........	..T......A	.C.....A..
ASN1	....	.A.....T..	G.........	..T......A	.C.....A..
ASN2	....	.A.....T..	G.........	..T......A	.C.....A..
ASN3	....	.A.....T..	G.........	..T......A	.C.....A..
ASN6	....	.A.....T..	G.........	..T......A	.C.....A..
ASN7	....	.A.....T..	G.........	..T......A	.C.....A..
AUS1	....	.A.....T..	G.........	..T......A	.C.....A..
AUS3	....	.A.....T..	G.........	..T......A	.C.....A..
AUS4	....	.A.....T..	G.........	..T......A	.C.....A..
AUS5	....	.A.....T..	G.........	..T......A	.C.....A..
AUS6	....	.A.....T..	G.........	..T......A	.C.....A..
Hs-ref	....	.A.....T..	G.........	..T......A	.C.....A..
Bonobo	....	.A........	C	....	.C.....A..
A-175	....	....T.....	.........G	....	.C.....A..
A-176	....	.A..T.....	.........G	....	.C.....A..
A-281	.....T....	.A........	.........G	....	.C.....A..
A-208	.....T....	.A........	.........G	....	.C.....A..
A-231	.....T....	.A........	.........G	....	.C.....A..
A-285	.....T....	.A........	.........G	....	.C.....A..
A-199	.....T....	.A........	.........G	....	.C.....A..
A-241	.....T....	.A........	.........G	....	.C.....A..
A-292	.....T....	.A........	.........G	....	.C.....A..
A-50	.....T....	.A........	.........G	....	.C.....A..
A-179	.....T....	.A........	.........G	....	.C.....A..
A-283	.....T....	.A........	.........G	....	.C.....A..
A-290	.....T....	.A........	.........G	....	.C.....A..
A-242	.....T....	.A........	.........G	....	.C.....A..
A-182	.....T....	.A........	.........G	.....G....	.C.....A..
A-288	.....T....	.A........	.........G	.....G....	.C.....A..
Pt-ref	.....T....	.A........	.........G	.....G....	.C.....A..
A-291	.....T....	.A........	.........G	.....G....	.C..G..A..
A-282	.....T....	.A........	.........G	.....G....	.C..G..A..
A-287	.....T....	.A........	.........G	.....G....	.C..G..A..
A-235	.....T....	.A........	.........G	.....G....	.C..G..A..

```
                15247       15257       15267       15277       15287
                |           |           |           |           |
Gorilla         TTACTCAGTA GATAGCCCTA CCCTTACACG ATTCTTTACC TTCCACTTTA
AFR1            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AFR2            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AFR6            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AFR7            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AFR8            C.......... ..C..T..C. ....C..... .......... ..T.....C.
EUR1            C.......... ..C..T..C. ....C..... .......... ..T.....C.
EUR2            C.......... ..C..T..C. ....C..... .......... ..T.....C.
EUR3            C.......... ..C..T..C. ....C..... .......... ..T.....C.
EUR4            C.......... ..C..T..C. ....C..... .......... ..T.....C.
EUR5            C.......... ..C..T..C. ....C..... .......... ..T.....C.
ASN1            C.......... ..C..T..C. ....C..... .......... ..T.....C.
ASN2            C.......... ..C..T..C. ....C..... .......... ..T.....C.
ASN3            C.......... ..C..T..C. ....C..... .......... ..T.....C.
ASN6            C.......... ..C..T..C. ....C..... .......... ..T.....C.
ASN7            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AUS1            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AUS3            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AUS4            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AUS5            C.......... ..C..T..C. ....C..... .......... ..T.....C.
AUS6            C.......... ..C..T..C. ....C..... .......... ..T.....C.
Hs-ref          C.......... ..C..T..C. ....C..... .......... ..T.....C.
Bonobo          C.......... ..C....... .......... ......C... C.........
A-175           C.......... ..C....... ....C..... ......C... ........C.
A-176           C.......... ..C....... .......... ......C... ..........
A-281           C.......... ..C....... .......... ......C... ..........
A-208           C.......... ..C....... .......... ......C... ..........
A-231           C.......... ..C....... .......... ......C... ..........
A-285           C.......... ..C....... .......... ......C... ..........
A-199           C.......... ..C....... .......... ......C... ..........
A-241           C.......... ..C....... .......... ......C... ..........
A-292           C.......... ..C....... .......... ......C... ..........
A-50            C.......... ..C....... .......... ......C... ..........
A-179           C.......... ..C....... .......... ......C... ..........
A-283           C.......... ..C....... .......... ......C... ..........
A-290           C.......... ..C....... .......... ......C... ..........
A-242           C.......... ..C....... .......... ......C... ..........
A-182           C.......... ..C....... .......... ......C... ..........
A-288           C.......... ..C....... .......... ......C... ..........
Pt-ref          C.......... ..C....... .......... ......C... ..........
A-291           C.......... ..C....... .......... ......C..T ..........
A-282           C.......... ..C....... .......... ......C..T ..........
A-287           C.......... ..C....... .......... ......C..T ..........
A-235           C.......... ..C....... .......... ......C..T ..........
```

```
                15297       15307       15317       15327       15337
                |           |           |           |           |
Gorilla         TCCTACCCTT  CATCATCACA  GCCCTAACAA  CCCTCCATCT  CCTATTTCTA
AFR1            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AFR2            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AFR6            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AFR7            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AFR8            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
EUR1            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
EUR2            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
EUR3            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
EUR4            ..T.G.....  ...T..TG..  ......G...  .A.....C..  ......CT.G
EUR5            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
ASN1            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
ASN2            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
ASN3            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
ASN6            ..T.G.....  ...T..T...  ......G..G  .A.....C..  ......CT.G
ASN7            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AUS1            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AUS3            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AUS4            ..T.G.....  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AUS5            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
AUS6            ..T.......  ...T..TG..  ......G..G  .A.....C..  ......CT.G
Hs-ref          ..T.G.....  ...T..TG..  ......G...  .A.....C..  ......CT.G
Bonobo          ..........  ...T......  ..........  .A..T.....  ......CT..
A-175           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-176           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-281           ..T....T..  ...T......  ..........  .A..T.....  ......CT..
A-208           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-231           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-285           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-199           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-241           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-292           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-50            ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-179           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-283           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-290           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-242           ..T.......  ...T......  ..........  .A..T.....  ......CT..
A-182           ..T.......  ..........  ..........  .A..T.....  ......CT..
A-288           ..T.......  ..........  ..........  .A..T.....  ......CT..
Pt-ref          ..T.......  ..........  ..........  .A..T.....  ......CT..
A-291           ..T.......  ..........  ..........  .A..T.....  ......CT..
A-282           ..T.......  ..........  ..........  .A..T.....  ......CT..
A-287           ..T.......  ..........  ..........  .A..T.....  ......CT..
A-235           ..T.......  ..........  ..........  .A..T.....  ......CT..
```

	15347	15357	15367	15377	15387
Gorilla	CACGAAACAG	GATCAAACAA	CCCTCTAGGC	ATCCCCTCCC	ACTCTGACAA
AFR1	........G.	....	...C.....A	...A......	.T..C..T..
AFR2	........G.	....	...C.....A	...A......	.T..C..T..
AFR6	........G.	....	...C.....A	...A......	.T..C..T..
AFR7	........G.	....	...C.....A	...A......	.T..C..T..
AFR8	........G.	....	...C.....A	...A......	.T..C..T..
EUR1	........G.	....	...C.....A	...A......	.T..C..T..
EUR2	........G.	....	...C.....A	...A......	.T..C..T..
EUR3	........G.	....	...C.....A	...A......	.T..C..T..
EUR4	........G.	....	...C.....A	...A......	.T..C..T..
EUR5	........G.	....	...C.....A	...A......	.T..C..T..
ASN1	........G.	....	...C.....A	...A......	.T..C..T..
ASN2	........G.	....	...C.....A	...A......	.T..C..T..
ASN3	........G.	....	...C.....A	...A......	.T..C..T..
ASN6	........G.	....	...C.....A	...A......	.T..C..T..
ASN7	........G.	....	...C.....A	...A......	.T..C..T..
AUS1	........G.	....	...C.....A	...A......	.T..C..T..
AUS3	........G.	....	...C.....A	...A......	.T..C..T..
AUS4	........G.	....	...C.....A	...A......	.T..C..T..
AUS5	........G.	....	...C.....A	...A......	.T..C..T..
AUS6	........G.	....	...C.....A	...A......	.T..C..T..
Hs-ref	........G.	....	...C.....A	...A......	.T..C..T..
Bonobo	....	.......T..	...C.....A	...A......	....C.....
A-175	....	.......T..	...C.....A	...A......	....C.....
A-176	....	.......T..	...C.....A	...A......	....C.....
A-281	....	.......T..	...C.....A	...A......	....C.....
A-208	....	.......T..	...C..G..A	...A......	....C.....
A-231	....	.......T..	...C..G..A	...A......	....C.....
A-285	....	.......T..	...C..G..A	...A......	....C.....
A-199	....	.......T..	...C..G..A	...A......	....C.....
A-241	....	.......T..	...C..G..A	...A......	....C.....
A-292	....	.......T..	...C..G..A	...A......	....C.....
A-50	....	.......T..	...C..G..A	...A......	....C.....
A-179	....	.......T..	...C..G..A	...A......	....C.....
A-283	....	.......T..	...C..G..A	...A......	....C.....
A-290	....	.......T..	...C..G..A	...A......	....C.....
A-242	....	.......T..	...C..G..A	...A......	....C.....
A-182	....	.......T..	...C.....A	...A......	....C.....
A-288	....	.......T..	...C.....A	...A......	....C.....
Pt-ref	....	.......T..	...C.....A	...A......	....C.....
A-291	....	.......T..	...C.....A	...A......	....C.....
A-282	....	.......T..	...C.....A	...A......	....C.....
A-287	....	.......T..	...C.....A	...A......	....C.....
A-235	....	.......T..	...C.....A	...A......	....C.....

	15397	15407	15417	15427	15437
	|	|	|	|	|
Gorilla	AATCACCTTC	CACCCCTACT	ACACAATCAA	AGACATCCTA	GGCCTATTCC
AFR1	....	.....T....	....	....GC...C	...T..C.T.
AFR2	....	.....T....	....	....GC...C	...T..C.T.
AFR6	....	.....T....	....	....GC...C	...T..C.T.
AFR7	....	.....T....	....	....GC...C	...T..C.T.
AFR8	....	.....T....	....	....GC...C	...T..C.T.
EUR1	....	.....T....	....	....GC...C	...T..C.T.
EUR2	....	.....T....	....	....GC...C	...T..C.T.
EUR3	....	.....T....	....	....GC...C	...T..C.T.
EUR4	....	.....T....	....	....GC...C	...T..C.T.
EUR5	....	.....T....	....	....GC...C	...T..C.T.
ASN1	....	.....T....	....	....GC...C	...T..C.T.
ASN2	....	.....T....	....	....GC...C	...T..C.T.
ASN3	....	.....T....	....	....GC...C	...T..C.T.
ASN6	....	.....T....	....	....GC...C	...T..C.T.
ASN7	....	.....T....	....	....GC...C	...T..C.T.
AUS1	....	.....T....	....	....GC...C	...T..C.T.
AUS3	....	.....T....	....	....GC...C	...T..C.T.
AUS4	....	.....T....	....	....GC...C	...T..C.T.
AUS5	....	.....T....	....	....GC...C	...T..C.T.
AUS6	....	.....T....	....	....GC...C	...T..C.T.
Hs-ref	....	.....T....	....	....GC...C	...T..C.T.
Bonobo	...T......	....	......C...	...T.....T	..TT......
A-175	...T......	....	....	...T.....T	..TT......
A-176	...T......	....	....	...T.....T	..TA......
A-281	...T......	....	....	...T.....T	...T......
A-208	...T......	....	....	.........T	...T......
A-231	...T......	....	....	.........T	...T......
A-285	...T......	....	....	.........T	...T......
A-199	...T......	....	....	...T.....T	...T......
A-241	...T......	....	....	...T.....T	...T......
A-292	...T......	....	....	...T.....T	...T......
A-50	...T......	....	....	...T.....T	...T......
A-179	...T......	....	....	...T.....T	...T......
A-283	...T......	....	....	...T.....T	...T......
A-290	...T......	....	....	...T.....T	...T......
A-242	...T......	....	....	...T.....T	...T......
A-182	...T......	....	....	...T.....T	...T......
A-288	...T......	....	....	...T.....T	...T......
Pt-ref	...T......	....	....	...T.....T	...T......
A-291	...T......	....	....	...T.....T	...T......
A-282	...T......	....	....	...T.....T	...T......
A-287	...T......	....	....	...T.....T	...T......
A-235	...T......	....	....	...T.....T	...T......

```
                15447       15457       15467       15477       15487
                |           |           |           |           |
Gorilla         TCTTTCTCCT  GACCTTGATA  ACATTAACAC  TATTCTCACC  AGACCTCCTA
AFR1            ....C..T..  CT....A..G  ..........  ..........  ..........
AFR2            ....C..T..  CT....A...  ..........  ..........  ..........
AFR6            ....C..T..  CT....A..G  ..........  ..........  ..........
AFR7            ....C..T..  CT....A..G  ..........  ..........  ..........
AFR8            ....C..T..  CT....A..G  ..........  ..........  ..........
EUR1            ....CA.T..  CT....A..G  ..........  ..........  ..........
EUR2            ....CA.T..  CT....A..G  ..........  ..........  .......T..
EUR3            ....C..T..  CT....A..G  ..........  ..........  ..........
EUR4            ....C..T..  CT....A..G  ..........  ..........  ..........
EUR5            ....CA.T..  CT....A..G  ..........  ..........  ..........
ASN1            ....C..T..  CT....A..G  ..........  ..........  ..........
ASN2            ....C..T..  CT....A..G  ..........  ..........  ..........
ASN3            ....C..T..  CT....A..G  ..........  ..........  ..........
ASN6            ....C..T..  CT....A..G  ..........  ..........  ..........
ASN7            ....C..T..  CT....A..G  ..........  ..........  ..........
AUS1            ....C..T..  CT....A..G  ..........  ..........  ..........
AUS3            ....C..T..  CT....A..G  ..........  ..........  ..........
AUS4            ....C..T..  CT....A..G  ..........  ..........  ..........
AUS5            ....C..T..  CT....A..G  ..........  ..........  ..........
AUS6            ....C..T..  CT....A..G  ..........  ..........  ..........
Hs-ref          ....C..T..  CT....A..G  ..........  ..........  ..........
Bonobo          .T..C.....  CG..C.A...  GT........  ....T.....  ..........
A-175           .T..C.....  T.T.C.A..G  ..........  ..........  ..........
A-176           .T..C.....  T.T.C.A..G  ..........  ..........  ..........
A-281           .T..C.....  T.T.C.A..G  ........G.  ..........  ..........
A-208           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-231           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-285           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-199           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-241           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-292           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-50            .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-179           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-283           .T..C.....  T.T.C.A..G  ..........  ..........  ....?....G
A-290           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-242           .T..C.....  T.T.C.A..G  ..........  ..........  .........G
A-182           ....C.....  T.T.C.A..G  ..........  ..........  ..........
A-288           .T..C.....  T.T.C.A..G  ..........  ..........  ..........
Pt-ref          .T..C.....  T.T.C.A..G  ..........  ..........  ..G.......
A-291           .T..C.....  T.T.C.A..G  ..........  ..........  ..........
A-282           .T..C.....  T.T.C.A..G  ..........  ..........  ..........
A-287           .T..C.....  T.T.C.A..G  ..........  ..........  ..........
A-235           .T..C.....  T.T.C.A..G  ..........  ..........  ..........
```

	15497	15507	15517	15527	15537
Gorilla	GGAGACCCAG	ACAACTACAC	CTTAGCCAAC	CCCCTAAGCA	CCCCACCCCA
AFR1	..C.....	....T..T..	.C.....	...T...A..	....T.....
AFR2	..C.....	....T..T..	.C.....	...T...A..	....T.....
AFR6	..C.....	....T..T..	.C.....	...T...A..	....T.....
AFR7	..C.....	....T..T..	.C.....	...T...A..	....T.....
AFR8	..C.....	....T..T..	.C.....	...T...A..	....T.....
EUR1	..C.....	....T..T..	.C.....	...T...A..	....T.....
EUR2	..C.....	....T..T..	.C.....	...T...A..	....T.....
EUR3	..C.....	....T..T..	.C.....	...T...A..	....T.....
EUR4	..C.....	....T..T..	.C.....	...T...A..	....T.....
EUR5	..C.....	....T..T..	.C.....	...T...A..	....T.....
ASN1	..C.....	....T..T..	.C.....	...T...A..	....T.....
ASN2	..C.....	....T..T..	.C.....	...T...A..	....T.....
ASN3	..C.....	....T..T..	.C.....	...T...A..	....T.....
ASN6	..C.....	....T..T..	.C.....	...T...A..	....T.....
ASN7	..C.....	....T..T..	.C.....	...T...A..	....T.....
AUS1	..C.....	....T..T..	.C.....	...T...A..	....T.....
AUS3	..C.....	....T..T..	.C.....	...T...A..	....T.....
AUS4	..C.....	....T..T..	.C.....	...T...A..	....T.....
AUS5	..C.....	....T..T..	.C.....	...T...A..	....T.....
AUS6	..C.....	....T..T..	.C.....	...T...A..	....T.....
Hs-ref	..C.....	....T..T..	.C.....	...T...A..	....T.....
Bonobo	..C..T....	....T.....	.C.....	.......T..	....
A-175	..C..T....	....	.C....T...	.......A..	....
A-176	..C..T....	....	.C....T...	.......A..	....
A-281	..C..T....	.......T..	.C....T...	...T...A..	....
A-208	..C..T....	.......T..	.C....T...	.......A..	....
A-231	..C..T....	.......T..	.C....T...	.......A..	....
A-285	..C..T....	.......T..	.?....T...	.......A..	....
A-199	..C..T....	.......T..	.C....T...	.......A..	....
A-241	..C..T....	.......T..	.C....T...	.......A..	....
A-292	..C..T....	.......T..	.C....T...	.......A..	....
A-50	..C..T....	.......T..	.C....T...	.......A..	....
A-179	..C..T....	.......T..	.C....T...	.......A..	....
A-283	..C..T....	.......T..	.C....T...	.......A..	....
A-290	..C..T....	.......T..	.C....T...	.......A..	....
A-242	..C..T....	.......T..	.C....T...	.......A..	....
A-182	..C..T....	.......T..	.C....T...	.......A..	....
A-288	..C..T....	.......T..	.C....T...	.......A..	....
Pt-ref	..C..T....	.......T..	.C....T...	.......A..	....
A-291	..T..T....	....	.C....T...	.......A..	....
A-282	..C..T....	....	.C....T...	.......A..	....
A-287	..C..T....	....	.C....T...	.......A..	....
A-235	..C..T....	....	.C....T...	.......A..	....

```
                15547       15557       15567       15577       15587
                |           |           |           |           |
Gorilla         CATCAAACCC  GAATGATATT  TCCTATTTGC  CTACGCAATT  CTCCGATCTG
AFR1            ......G...  ..........  .......C..  ....A.....  ........C.
AFR2            ......G...  ..........  .......C..  ....A.....  ........C.
AFR6            ......G...  ..........  .......C..  ....A.....  ........C.
AFR7            ......G...  ..........  .......C..  ....A.....  ........C.
AFR8            ......G...  ..........  .......C..  ....A.....  ........C.
EUR1            ......G...  ..........  .......C..  ....A.....  ........C.
EUR2            ......G...  ..........  .......C..  ....A.....  ........C.
EUR3            ......G...  ..........  .......C..  ....A.....  ........C.
EUR4            ......G...  ..........  .......C..  ....A.....  ........C.
EUR5            ......G...  ..........  .......C..  ....A.....  ........C.
ASN1            ......G...  ..........  .......C..  ....A.....  ........C.
ASN2            ......G...  ..........  .......C..  ....A.....  ........C.
ASN3            ......G...  ..........  .......C..  ....A.....  ........C.
ASN6            ......G...  ..........  .......C..  ....A.....  ........C.
ASN7            ......G...  ..........  .......C..  ....A.....  ........C.
AUS1            ......G...  ..........  .......C..  ....A.....  ........C.
AUS3            ......G...  ..........  .......C..  ....A.....  ........C.
AUS4            ......G...  ..........  .......C..  ....A.....  ........C.
AUS5            ......G...  ..........  .......C..  ....A.....  ........C.
AUS6            ......G...  ..........  .......C..  ....A.....  ........C.
Hs-ref          ......G...  ..........  .......C..  ....A.....  ........C.
Bonobo          ...T..G...  ..G..G....  .T........  ....A.....  ........C.
A-175           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-176           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-281           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-208           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-231           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-285           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-199           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-241           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-292           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-50            ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-179           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-283           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-290           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-242           ...T......  ..G.....C.  .T........  ....A.....  ........CA
A-182           ...T......  ..G.....C.  .T........  ....A....C  ........CA
A-288           ...T......  ..G.....C.  .T........  ....A....C  ........CA
Pt-ref          ...T......  ..G.....C.  .T........  ....A....C  ........CA
A-291           ...T......  ..G.....C.  .T........  ....A....C  ........CA
A-282           ...T......  ..G.....C.  .T........  ....A....C  ........CA
A-287           ...T......  ..G.....C.  .T........  ....A....C  ........CA
A-235           ...T......  ..G.....C.  .T........  ....A....C  ........CA
```

```
                15597       15607       15617       15627       15637
                |           |           |           |           |
Gorilla         TCCCCAATAA  ACTAGGAGGC  GTCTTAGCTC  TATTACTATC  TATTCTCATC
AFR1            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AFR2            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AFR6            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AFR7            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AFR8            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
EUR1            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
EUR2            ....T..C..  G.........  ...C.T..C.  ..........  C..C......
EUR3            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
EUR4            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
EUR5            ....T..C..  G.........  ...C.T..C.  ..........  C..C......
ASN1            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
ASN2            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
ASN3            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
ASN6            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
ASN7            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AUS1            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AUS3            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AUS4            ....T..C..  G.........  ...C.T..C.  ..........  C..C......
AUS5            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
AUS6            ....T..C..  ..........  ...C.T..C.  ..........  C..C......
Hs-ref          ....T..C..  ..........  ...C.T..C.  ..........  C..C......
Bonobo          .......C..  .........T  ..TC.C..C.  ..C.......  ...C......
A-175           .......C..  .........T  ...C.C..CT  ..C.G.....  ...C..G...
A-176           .......C..  .........T  ...C.C..CT  ..C.G.....  ...C..G...
A-281           .......C..  ..........  ...C.C..C.  ..C.T.....  ...C..G...
A-208           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-231           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-285           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-199           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-241           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-292           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-50            .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-179           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-283           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-290           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-242           .......C..  ..........  ...C.C..C.  ..C.......  ...C..G...
A-182           .......C..  ..........  ...C.C..C.  ..C.......  ...C..A...
A-288           .......C..  ..........  ...C.C..C.  ..C.......  ...C..A...
Pt-ref          .......C..  ..........  ...C.C..C.  ..C.......  ...C..A...
A-291           .......C..  ..........  ...C.C..C.  ..C.......  ...C..A...
A-282           .......C..  ..........  ...C.C..C.  ..C.......  ...C..A...
A-287           .......C..  ..........  ...C.C..C.  ..C.......  ...C..A...
A-235           .......C..  ..........  ...C.C..C.  ..C.......  ...C..A...
```

```
                15647      15657      15667      15677      15687
                |          |          |          |          |
Gorilla         CTAGCAATAA TTCCTATTCT CCACATATCC AAACAACAAA GCATAATATT
AFR1            .......... .C..C..C.. ...T...... .......... ..........
AFR2            .......... .C..C..C.. ...T...... .......... .....G....
AFR6            .......... .C..C..C.. ...T...... .......... ..........
AFR7            .......... .C..C..C.. ...T...... .......... ..........
AFR8            .......... .C..C..C.. .......... .......... ..........
EUR1            .......... .C..C..C.. ...T...... .......... ..........
EUR2            .......... .C..C..C.. ...T...... .......... ..........
EUR3            .......... .C..C..C.. ...T...... .......... ..........
EUR4            .......... .C..C..C.. ...T...... .......... ..........
EUR5            .......... .C..C..C.. ...T...... .......... ..........
ASN1            .......... .C..C..C.. ...T...... .......... ..........
ASN2            .......... .C..C..C.. ...T...... .......... ..........
ASN3            .......... .C..C..C.. ...T...... .......... ..........
ASN6            .......... .C..C..C.. ...T...... .......... ..........
ASN7            .......... .C..C..C.. ...T...... .......... ..........
AUS1            .......... .C..C..C.. ...T...... .......... ..........
AUS3            .......... .C..C..C.. ...T...... .......... ..........
AUS4            .......... .C..C..C.. .......... .......... ..........
AUS5            .......C.. .C..C..C.. ...T...... .......... ..........
AUS6            .......... .C..C..C.. ...T...... .......... ..........
Hs-ref          .......... .C..C..C.. ...T...... .......... ..........
Bonobo          ......G... .C.....C.. .....C.... .......... ..........
A-175           ......GC.. .C...G.C.. ...T.C.... .......... ..........
A-176           ......GC.. .C...G.C.. ...T.C.... .......... ..........
A-281           ......GC.. .C...G.C.. .....C.... .......... ..........
A-208           ......GC.. .C...G.C.. .....C.... .......... ..........
A-231           ......GC.. .C...G.C.. .....C.... .......... ..........
A-285           ......GC.. .C...G.C.. .....C.... .......... ..........
A-199           ......GC.. .C..CG.C.. .....?.... .......... ..........
A-241           ......GC.. .C...G.C.. .....C.... .......... ..........
A-292           ......GC.. .C...G.C.. .....C.... .......... ..........
A-50            ......GC.. .C..CG.C.. .....C.... .......... ..........
A-179           ......GC.. .C...G.C.. .....C.... .......... ..........
A-283           ......GC.. .C...G.C.. .....C.... .......... ..........
A-290           ......GC.. .C...G.C.. .....C.... .......... ..........
A-242           ......GC.. .C...G.C.. .....C.... .......... ..........
A-182           ......GC.. .....G.C.. .....C.... .......... ..........
A-288           ...A..GC.. .C...G.C.. .....C.... .......... ..........
Pt-ref          ...A..GC.. .C...G.C.. .....C.... .......... ..........
A-291           ...A..GC.. .C...G.C.. .....C.... .......... ..........
A-282           ...A..GC.. .C...G.C.. .....C.... .......... ..........
A-287           ...A..GC.. .C...G.C.. .....C.... .......... ..........
A-235           ...A..GC.. .C...G.C.. .....C.... .......... ..........
```

	15697	15707	15717	15727	15737
	|	|	|	|	|
Gorilla	CCGCCCATTA	AGCCAACTAC	TCTACTGATT	CCTAATCGCA	GACCTCTTCA
AFR1	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AFR2	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AFR6	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AFR7	T.....C..	.....TC..	.T.....C.	....GC....	.....C...
AFR8	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
EUR1	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
EUR2	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
EUR3	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
EUR4	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
EUR5	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
ASN1	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
ASN2	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
ASN3	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
ASN6	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
ASN7	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AUS1	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AUS3	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AUS4	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AUS5	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
AUS6	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
Hs-ref	T.....C..	.....TC..	.T..T...C.	....GC....	.....C...
Bonobo	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-175	.....C..	........G.	.T.....C.	....GC....	.....C.T.
A-176	.....C..	........G.	.T.....C.	....GC....	.....C.T.
A-281	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-208	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-231	T.....C..	........G.	.T.....C.	....GC.A..	.....TC...
A-285	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-199	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-241	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-292	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-50	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-179	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-283	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-290	T.....C..	........G.	.T.....C.	....GC.A..	.....C.T.
A-242	T.....C..	...?....G.	.T.....C.	....GC.A..	.....C.T.
A-182	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-288	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
Pt-ref	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-291	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-282	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-287	T.....C..	........G.	.T.....C.	....GC.A..	.....C...
A-235	T.....C..	........G.	.T.....C.	....GC.A..	.....C...

```
                15747       15757       15767       15777       15787
                |           |           |           |           |
Gorilla         CCCTAACCTG AATCGGAGGA CAACCAGTAA GCTACCCCTT CATCACCATT
AFR1            TT......... .......... .......... .......T.. T.C..T....
AFR2            TT......... .......... .......... .......T.. T.C..T....
AFR6            TT......... .......... .......... .......T.. T.C..T....
AFR7            TT......... .......... .......... .......T.. T.C..T....
AFR8            TT......... .......... .......... .......T.. T.C..T....
EUR1            TT......... .......... .......... .......T.. T.C..T....
EUR2            TT......... .......... .......... .......T.. T.C..T....
EUR3            TT......... .......... .......... .......T.. T.C..T....
EUR4            TT......... .......... .......... .......T.. T.C..T....
EUR5            TT......... .......... .......... .......T.. T.C..T....
ASN1            TT......... .......... .......... .......T.. T.C..T....
ASN2            TT......... .......... .......... .......T.. T.C..T....
ASN3            TT......... .......... .......... .......T.. T.C..T....
ASN6            TT......... .......... .......... .......T.. T.C..T....
ASN7            TT......... .......... .......... .......T.. T.C..T....
AUS1            TT......... .......... .......... .......T.. T.C..T....
AUS3            TT......... .......... .......... .......T.. T.C..T....
AUS4            TT......... .......... .......... .......T.. T.C..T....
AUS5            TT......... .......... .......... .......T.. T.C..T....
AUS6            TT......... .......... .......... .......T.. T.C..T....
Hs-ref          TT......... .......... .......... .......T.. T.C..T....
Bonobo          T........... .......... .......... .......... .........C
A-175           T........... .........G .......... .......... .........C
A-176           T........... .........G .......... .......... .........C
A-281           T.........  .G........ .......... .......... .........C
A-208           T........... .......... .......... .......... .........C
A-231           T........... .......... .......... .......... .........C
A-285           T........... .......... .......... .......... .........C
A-199           T........... .......... .......... .......... .........C
A-241           T........... .......... .......... .......... .........C
A-292           T........... .......... .......... .......... .........C
A-50            T........... .......... .......... .......... .........C
A-179           T........... .......... .......... .......... .........C
A-283           T........... .......... .......... .......... .........C
A-290           T........... .......... .......... .......... .........C
A-242           T........... .......... .......... .......... .........C
A-182           T........... .......... .......... .......... .........C
A-288           T........... .......... .......... .......... .........C
Pt-ref          T........... .......... .......... .......... .........C
A-291           T........... .......... .......... .......... .........C
A-282           T........... .......... .......... .......... .........C
A-287           T........... .......... .......... .......... .........C
A-235           T........... .......... .......... .......... .........C
```

	15797	15807	15817	15827	15837
Gorilla	GGGCAAGTAG	CATCCGTACT	ATACTTCACG	ACAATCCTAT	TCCTGATACC
AFR1	..A.....	.....	.....A	.....A	....A.....
AFR2	..A.....	.....	.....A	.....A	....A.....
AFR6	..A.....	.....	.....A	.....A	....A.....
AFR7	..A.....	.....	.....A	.....A	....A.....
AFR8	..A.....	.....	.....A	.....A	....A.....
EUR1	..A.....	.....	.....A	.....A	....A.....
EUR2	..A.....	.....	.....A	.....A	....A.....
EUR3	..A.....	.....	.....A	.....A	....A.....
EUR4	..A.....	.....	.....A	.....A	....A.....
EUR5	..A.....	.....	.....A	.....A	....A.....
ASN1	..A.....	.....	.....A	.....A	....A.....
ASN2	..A.....	.....	.....A	.....A	....A.....
ASN3	..A.....	.....	.....A	.....A	....A.....
ASN6	..A.....	.....	.....A	.....A	....A.....
ASN7	..A.....	.....	.....A	.....A	....A.....
AUS1	..A.....	.....	.....A	.....A	....A.....
AUS3	..A.....	.....	.....A	.....A	....A.....
AUS4	..A.....	.....	G.....A	.....A	....A.....
AUS5	..A.....	.....	.....A	.....A	....A.....
AUS6	..A.....	.....	.....A	.....A	....A.....
Hs-ref	..A.....	.....	.....A	.....A	....A.....
Bonobo	.....	.....T.	G.....A	.....A	....A.....
A-175	..A.....	.....T.	.....A	.....A	....A.....
A-176	..A.....	.....T.	.....	.....A	....A.....
A-281	..A...A...	.....T.	.....A	.....A	....A.....
A-208	..A...A...	.....T.	.....A	.....A	....A.....
A-231	..A...A...	.....T.	.....A	.....A	....A.....
A-285	..A...A...	.....T.	.....A	.....A	....A.....
A-199	..A...A...	.....T.	.....A	.....A	....A.....
A-241	..A...A...	.....T.	.....A	.....A	....A.....
A-292	..A...A...	.....T.	.....A	.....A	....A.....
A-50	..A...A...	.....T.	.....A	.....A	....A.....
A-179	..A...A...	.....T.	.....A	.....A	....A.....
A-283	..A...A...	.....T.	.....A	.....A	....A.....
A-290	..A...A...	.....T.	.....A	.....A	....A.....
A-242	..A...A...	.....T.	.....A	.....?A	....A.....
A-182	..A...A...	.....T.	.....A	.....A	....A.....
A-288	..A...A...	.....T.	.....A	.....A	....A.....
Pt-ref	..A...A...	.....T.	.....A	.....A	....A.....
A-291	..A...A...	.....T.	.....A	.....A	....A.....
A-282	..A...A...	.....T.	.....A	.....A	....A.....
A-287	..A...A...	.....T.	.....A	.....A	....A.....
A-235	..A...A...	.....T.	.....A	.....A	....A.....

	15847	15857	15867	15877
	|	|	|	|
Gorilla	AATCACATCC	CTGATCGAAA	ACAAAATACT	CAAATGAACC
AFR1	..CT.TC...	..A..T....	....	......GG..
AFR2	..CT.TC...	..A..T....	....	......GG..
AFR6	..CT.TC...	..A..T....	....	......GG..
AFR7	..CT.TC...	..A..T....	....	......GG..
AFR8	..CT.TC...	..A..T....	....	......GG..
EUR1	..CT.TC...	..A..T....	....	......GG..
EUR2	..CT.TC...	..A..T....	....	......GG..
EUR3	..CT.TC...	..A..T....	....	......GG..
EUR4	..CT.TC...	..A..T....	....	......GG..
EUR5	..CT.TC...	..A..T....	....	......GG..
ASN1	..CT.TC...	..A..T....	....	......GG..
ASN2	..CT.TC...	..A..T....	....	......GG..
ASN3	..CT.TC...	..A..T....	....	......GG..
ASN6	..CT.TC...	..A..T....	....	.......G..
ASN7	..CT.TC...	..A..T....	....	......GG..
AUS1	..CT.TC...	..A..T....	....	......GG..
AUS3	..CT.TC...	..A..T....	....	......GG..
AUS4	..CT.TC...	..A..T....	....	......GG..
AUS5	..CT.TC...	..A..T....	....	......GG..
AUS6	..CT.TC...	..A..T....	....	......GG..
Hs-ref	..CT.TC...	..A..T....	....	......GG..
Bonobo	...T.TC...	..A.......	....	.G.....G..
A-175	...T..C...	..A.......	....	TG........
A-176	...T..C...	..A.......	....	TG........
A-281	....G.C...	..A.......	....	TG....G...
A-208	....G.C..T	..A.......	....	TG........
A-231	....G.C..T	..A.......	....	TG........
A-285	....G.C..T	..A.......	....	TG........
A-199	....G.C..T	..A.......	....	TG........
A-241	....G.C..T	..A.......	....	TG........
A-292	....G.C..T	..A.......	....	TG........
A-50	....G.C..T	..A.......	....	TG........
A-179	....G.C..T	..A.......	....	TG........
A-283	....G.C..T	..A.......	....	TG........
A-290	....G.C..T	..A.......	....	TG........
A-242	....G.C..T	..A.......	....	TG........
A-182	....G.C..T	..A.......	....	TG........
A-288	....G.C..T	..A.......	....	TG........
Pt-ref	....G.C..T	..A.......	....	TG........
A-291	....G.C..T	..A.......	....	TG........
A-282	....G.C..G	..A.......	....	TG........
A-287	....G.C..G	..A.......	....	TG........
A-235	....G.C..G	..A.......	....	TG........

APPENDIX D

D1. UNAMBIGUOUS NUCLEOTIDE CHANGES ALONG BRANCHES ① TO ⑦ OF THE *ND2* NETWORK SHOWN IN FIGURE 4.3 *C*

Branch ①

Synonymous: 4475, 4520, 4523, 4532, 4616, 4634, 4688, 4736, 4739, 4760, 4778, 4823, 4844, 4847, 4883, 4904a, 4919, 4947, 4988, 5000, 5003, 5027, 5057, 5066, 5072, 5082, 5084, 5087, 5102, 5121, 5130, 5151, 5165, 5168, 5174, 5177, 5180, 5204, 5228, 5250, 5252, 5291, 5300, 5312, 5315, 5333, 5348, 5360, 5372, 5405, 5429, 5438, 5456, 5462, 5465, 5492, 5507, 5508

Nonsynonymous: 4554, 4726, 4734, 4764, 4767, 4886a, 4914, 4915, 4956, 5028, 5112, 5134, 5262, 5292, 5293, 5302, 5418, 5424, 5505

Branch ②

Synonymous: 4526, 4580, 4592, 4598, 4619, 4622, 4643, 4655, 4679, 4703, 4769, 4820, 4838, 4841, 4853, 4856, 4859, 4907, 4940, 4967, 4973, 4977, 4991, 5004, 5021, 5054, 5147, 5195, 5196, 5201, 5231, 5237, 5249, 5264, 5288, 5342, 5393, 5441, 5450, 5477, 5483, 5495

Nonsynonymous: 4491, 4696, 4843, 4908, 5263, 5320, 5460

Branch ③

Synonymous: 4481, 4553, 4577, 4646, 4649, 4667, 4676, 4763, 4775, 4892, 4916, 4937, 4949, 4955, 5033, 5096, 5120, 5192, 5261, 5267, 5327, 5387, 5396, 5417

Nonsynonymous: 4615, 4924

Branch ④

Synonymous: 4673, 4733, 4790, 5300, 5351, 5372, 5399, 5435, 5504

Nonsynonymous: 4521, 4531, 4768, 4904b, 5112, 5311

Branch ⑤

Synonymous: 4682, 4772, 4808, 4875, 4886b, 5048, 5093, 5117, 5375, 5474

Nonsynonymous: 4596, 4702

Branch ⑥

Synonymous: 4520, 4634, 4883, 5057, 5201, 5252

Nonsynonymous: 4512

Branch ⑦

Synonymous: 4835, 4847, 5231, 5333

Nonsynonymous: 4729, 4738, 4924, 5263, 5440

D2. UNAMBIGUOUS NUCLEOTIDE CHANGES ALONG BRANCHES ① TO ⑦ OF THE *cyt b* NETWORK SHOWN IN FIGURE 4.4 *C*

Branch ①

Synonymous: 14755a, 14783, 14803, 14812, 14818, 14860, 14867, 14890, 14917, 14920, 14944, 14962, 14992, 15022, 15031, 15046, 15055, 15106, 15184, 15193, 15208, 15238, 15244, 15247, 15259, 15310, 15328, 15343, 15344, 15370, 15376, 15391, 15440a, 15451, 15457a, 15463, 15499, 15518, 15595, 15604, 15620, 15625, 15640, 15658, 15664, 15697, 15704, 15718, 15826, 15841, 15850, 15859

Nonsynonymous: 14766, 14777, 14778, 14979, 15071, 15122, 15380, 15534, 15581, 15725, 15731, 15732, 15743, 15747, 15836

Either: 15853

Branch ②

Synonymous: 14773, 14869, 14875, 14905, 14935, 14956, 14968, 14998, 15001, 15061, 15097, 15109, 15115, 15148, 15163, 15214, 15217, 15229, 15262, 15265, 15271, 15289, 15295, 15301, 15313, 15334, 15346, 15355, 15388, 15394, 15412, 15445, 15454, 15514, 15530, 15541, 15574, 15601, 15637, 15661, 15670, 15721, 15748, 15784, 15787, 15862, 15883

Nonsynonymous: 15110, 15236, 15314, 15323, 15326, 15431, 15432, 15443, 15458, 15714, 15789, 15792, 15849

Either: 15074, 15713

Branch ③

Synonymous: 14791, 14800, 14827, 14854, 14857, 14902, 14941, 14974, 15016, 15091, 15104, 15130, 15133, 15191, 15196, 15283, 15331, 15364, 15400, 15430, 15439, 15448, 15461, 15502, 15550, 15559, 15568, 15616, 15629, 15715, 15745, 15796, 15815

Nonsynonymous: 14757, 15653, 15672, 15734, 15878

Branch ④

Synonymous: 14767, 14824, 14953, 15481, 15562, 15619, 15817

Nonsynonymous: 15014, 15038, 15287, 15423, 15458, 15467, 15468, 15534

Branch ⑤

Synonymous: 14839, 14891, 14938, 14977, 15457b, 15523, 15565, 15643a, 15877

Nonsynonymous: 15459, 15596, 15654, 15662

Branch ⑥

Synonymous: 14881, 14992, 15049, 15097, 15202, 15439, 15514, 15616, 15745, 15850

Nonsynonymous: 15803, 15851

Branch ⑦

Synonymous: 14902, 15001, 15196, 15211, 15626, 15631a, 15670, 15697, 15766

Nonsynonymous: 15734

REFERENCES

REFERENCES

Altheide, T. K. and M. F. Hammer. 1997. Evidence for a possible Asian origin of YAP^+ Y chromosomes. *Am. J. Hum. Genet.* **61:** 462-466.

Anderson, S., A. T. Bankier, B. G. Barrell *et al.* 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290:** 457-465.

Ankel-Simons, F. and J. M. Cummins. 1996. Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. *Proc. Natl. Acad. Sci. USA* **93:** 13859-13863.

Aquadro, C. F. and B. D. Greenberg. 1983. Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics* **103:** 287-312.

Armour, J. A. L., T. Anttinen, C. A. May, E. E. Vega, A. Sajantila, J. R. Kidd, K. K. Kidd, J. Bertranpetit, S. Pääbo and A. J. Jeffreys. 1996. Minisatellite diversity supports a recent African origin for modern humans. *Nat. Genet.* **13:** 154-160.

Attardi, G. 1985. Animal mitochondrial DNA: an extreme example of genetic economy. *Int. Rev. Cytol.* **93:** 93-145.

Attardi, G., A. Chomyn, R. F. Doolittle, P. Mariottini and C. I. Ragan. 1986. Seven unidentified reading frames of human mitochondrial DNA encode subunits of the respiratory chain NADH dehydrogenase. *Cold Spring Harb. Symp. Quant. Biol.* **51**: 103-114.

Avise, J. C. 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. *Phil. Trans. R. Soc. Lond. B* **312:** 325-342.

Avise, J. C. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annu. Rev. Genet.* **25:** 45-69.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18:** 489-522.

Ayala, F. J. 1995. The myth of eve: molecualr biology and human origins. *Science* **270:** 1930-1936.

Ayala, F. J. and A. A. Escalante. 1996. The evolution of human populations: a molecular perspective. *Mol. Phylogenet. Evol.* **5:** 188-201.

Ayala, F. J., A. Escalante, C. O'hUigin and J. Klein. 1994. Molecular genetics of speciation and human origins. *Proc. Natl. Acad. Sci. USA* **91:** 6787-6794.

Bailey, W. J., K. Hayasaka, C. G. Skinner, S. Kehoe, L. C. Sieu, J. L. Slightom and M. Goodman. 1992. Reexamination of the African hominoid trichotomy with additional sequences from the primate β-globin gene cluster. *Mol. Phylogenet. Evol.* **1:** 97-135.

Ballard, J. W. O. and M. Kreitman. 1994. Unraveling selection in the mitochondrial genome of *Drosophila*. *Genetics* **138:** 757-772.

Bandelt, H., P. Forster, B. C. Sykes and M. B. Richards. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141:** 743-753.

Batzer, M. A. and P. L. Deininger. 1991. A human-specific subfamily of *Alu* sequences. *Genomics* **9:** 481-487.

Batzer, M. A., V. A. Gudi, J. C. Mena, D. W. Foltz, R. J. Herrera and P. L. Deininger. 1991. Amplification dynamics of human-specific (HS) Alu family members. *Nuc. Acids Res.* **19:** 3619-3623.

Batzer, M. A., M. Stoneking, M. Alegria-Hartman *et al.* 1994. African origin of human-specific polymorphic *Alu* insertions. *Proc. Natl. Acad. Sci. USA* **91:** 12288-12292.

Batzer, M. A., S. S. Arcot, J. W. Phinney *et al.* 1996. Genetic variation of recent *Alu* insertions in human populations. *J. Mol. Evol.* **42:** 22-29.

Bermingham, E., T. Lamb and J. C. Avise. 1986. Size polymorphism and heteroplasmy in the mitochondrial DNA of lower vertebrates. *J. Hered.* **77:** 249-252.

Berry, A. J., J. W. Ajioka and M. Kreitman. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129:** 1111-1117.

Birky, C. W. 1983. Relaxed cellular controls and organelle heredity. *Science* **222:** 468-475.

Birky, C. W., T. Maruyama and P. A. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103:** 513-527.

Birky, C. W., P. A. Fuerst and T. Maruyama. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121:** 613-627.

Blanchard, J. L. and G. W. Schmidt. 1996. Mitochondrial DNA migration events in yeast and humans: Integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol. Biol. Evol.* **13:** 537-548.

Board, P. G., C. J. Gibbs and D. C. Gajdusek. 1981. Polymorphism of erythrocyte glyoxalase II in anthropoid primates. *Folia Primatologica* **36:** 138-143.

Bowcock, A. M., C. Bucci, J. M. Hebert, J. R. Kidd, K. K. Kidd, J. S. Friedlaender and L. L. Cavalli-Sforza. 1987. Study of 47 DNA markers in five populations from four continents. *Gene Geog.* **1:** 47-64.

Bowcock, A. M., J. M. Herbet, J. L. Mountain, J. R. Kidd, J. Rogers, K. K. Kidd and L. L. Cavalli-Sforza. 1991*a*. Study of an additional 58 DNA markers in five human populations from four continents. *Gene Geog.* **5:** 151-173.

Bowcock, A. M., J. R. Kidd, J. L. Mountain, J. M. Hebert, L. Carotenuto, K. K. Kidd and L. L. Cavalli-Sforza. 1991*b*. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* **88:** 839-843.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368:** 455-457.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140:** 783-796.

Breguet, G., R. Butler, E. Butler-Brunner and A. Sanchez-Mazas. 1990. A worldwide population study of the Ag-system haplotypes, a genetic polymorphism of human low-density lipoprotein. *Am. J. Hum. Genet.* **46:** 502-517.

Brookfield, J. F. Y. and P. M. Sharp. 1994. Neutralism and selectionism face up to DNA data. *Trends Genet.* **10:** 109-111.

Brown, W. M. 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. USA* **77:** 3605-3609.

Brown, W. M. 1983. Evolution of animal mitochondrial DNA. Pp. 62-88 in M. Nei and R. K. Koehn (eds). *Evolution of genes and proteins.* Sinauer, Sunderland, Massachusetts.

Brown, W. M., M. George and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **76:** 1967-1971.

Brown, W. M., E. M. Prager, A. Wang and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* **18:** 225-239.

Bruce, E. J. and F. J. Ayala. 1979. Phylogenetic relationships between man and the apes: electrophoretic evidence. *Evolution* **33:** 1040-1056.

Burrows, W. and O. A. Ryder. 1997. Y-chromosome variation in great apes. *Nature* **385:** 125-126.

Cann, R. L. and A. C. Wilson. 1983. Length mutations in human mitochondrial DNA. *Genetics* **104:** 699-711.

Cann, R. L., W. M. Brown and A. C. Wilson. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* **106:** 479-499.

Cann, R. L., M. Stoneking and A. C. Wilson. 1987*a*. Mitochondrial DNA and human evolution. *Nature* **325:** 31-36.

Cann, R. L., M. Stoneking and A. C. Wilson. 1987*b*. Reply to letter. *Nature* **329:** 111-112.

Cantatore, P. and G. Attardi. 1980. Mapping of nascent light and heavy transcripts on the physical map of HeLa cell mitochondrial DNA. *Nuc. Acids. Res.* **8:** 2605-2625.

Castelloe, J. and A. R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. *Mol. Phylogenet. Evol.* **3:** 102-113.

Cavalli-Sforza, L. L., P. Menozzi and A. Piazza. 1994. *The history and geography of human genes*. Princeton University Press, Princeton.

Cavalli-Sforza, L. L., A. Piazza, P. Menozzi and J. L. Mountain. 1988. Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* **85:** 6002-6006.

Chakraborty, R., M. Kimmel, D. N. Stivers, L. J. Davison and R. Deka. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA* **94:** 1041-1046.

Charlesworth, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63:** 213-227.

Charlesworth, B., M. T. Morgan and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134:** 1289-1303.

Charlesworth, D., B. Charlesworth and M. T. Morgan. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* **141:** 1619-1632.

Chen, L. Z., S. Easteal, P. G. Board and R. L. Kirk. 1990. Evolution of β-globin haplotypes in human populations. *Mol. Biol. Evol.* **7:** 423-437.

Chen, Y.-S., A. Torroni, L. Excoffier, A. S. Santachiara-Benerecetti and D. C. Wallace. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57:** 133-149.

Chin, A. N. 1991. Mitochondrial DNA sequence variation in southern Chinese. Thesis, The Australian National University.

Clayton, D. A. 1982. Replication of animal mitochondrial DNA. *Cell* **28:** 693-705.

Clayton, D. A., J. N. Doda and E. C. Friedberg. 1974. The absence of a pyrimidine dimer repair mechanism in mammalian mitochondria. *Proc. Natl. Acad. Sci. USA* **71:** 2777-2781.

Coon, C. S. 1962. *The origin of races.* Knopf, New York.

Craig, J., S. Fowler, L. A. Burgoyne, A. C. Scott and H. W. J. Harding. 1988. Repetitive deoxyribonucleic acid (DNA) and human genome variation: A concise review relevant to forensic biology. *J. Foren. Sci.* **33:** 1111-1126.

Crouau-Roy, B., S. Service, M. Slatkin and N. Freimer. 1996. A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. *Hum. Mol. Genet.* **5:** 1131-1137.

D'Andrade, R. and P. A. Morin. 1996. Chimpanzee and human mitochondrial DNA: a principle components and individual-by-site analysis. *Am. Anthropol.* **98:** 352-370.

Darlu, P. and P. Tassy. 1987. Disputed African origin of human populations. *Nature* **329:** 111.

Dawid, I. B. 1972. Evolution of mitochondrial DNA sequences in *Xenopus*. *Rev. Biochem.* **53:** 573-594.

Deininger, P. L. and M. A. Batzer. 1993. Evolution of retroposons. Pp. 157-196 in M. K. Hecht, R. J. MacIntyre and M. T. Clegg (eds). *Evolutionary biology*. Plenum-Press, New York.

Deka, R., M. D. Shriver, L. M. Yu, L. Jin, C. E. Aston, R. Chakraborty and R. E. Ferrell. 1994. Conservation of human chromosome 13 polymorphic microsatellite $(CA)_n$ repeats in chimpanzees. *Genomics* **22:** 226-230.

Deka, R., L. Jin, M. D. Shriver, L. M. Yu, S. DeCroo, J. Hundrieser, C. H. Bunker, R. E. Ferrell and R. Chakraborty. 1995. Population genetics and dinucleotide $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms in world populations. *Am. J. Hum. Genet.* **56**: 461-474.

DeSalle, R. and A. R. Templeton. 1988. Founder effects and the rate of mitochondrial DNA evolution in Hawaiian *Drosophila* . *Evolution* **42:** 1076-1084.

Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin and N. B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* **91:** 3166-3170.

Di Rienzo, A. and A. C. Wilson. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **88:** 1597-1601.

Donnelly, P., S. Tavaré, D. J. Balding and R. C. Griffiths. 1996. Estimating the age of the common ancestor of men from the *ZFY* intron. *Science* **272:** 1357-1359.

Dorit, R. L., H. Akashi and W. Gilbert. 1995. Absence of polymorphism at the *ZFY* locus on the human Y chromosome. *Science* **268:** 1183-1185.

Dowsett, A. P. and M. W. Young. 1982. Differing levels of dispersed repetitive DNA among closely related species of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **79:** 4570-4574.

Easteal, S., C. Collet and D. Betty. 1995. *The mammalian molecular clock.* Springer-Verlag, Austin.

Edwards, A., A. Civitello, H. A. Hammond and C. T. Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49:** 746-756.

Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12:** 241-253.

Ellegren, H., C. R. Primmer and B. C. Sheldon. 1995. Microsatellite 'evolution': directionality or bias? *Nat. Genet.* **11:** 360-362.

Ely, J., R. Deka, R. Chakraborty and R. E. Ferrell. 1992. Comparison of five tandem repeat loci between humans and chimpanzees. *Genomics* **14:** 692-698.

Erlich, H. A., T. F. Bergstrom, M. Stoneking and U. Gyllensten. 1996. HLA sequence polymorphism and the origin of humans. *Science* **274:** 1552-1554.

Esposti, M. D., S. De Vries, M. Crimi, A. Ghelli, T. Patarnello and A. Meyer. 1993. Mitochondrial cytochrome *b*: evolution and structure of the protein. *Biochimica et Biophysica Acta* **1143:** 243-271.

Excoffier, L. 1990. Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *J. Mol. Evol.* **30:** 125-139.

Excoffier, L. and A. Langaney. 1989. Origin and differentiation of human mitochondrial DNA. *Am. J. Hum. Genet.* **44:** 73-85.

Feldman, M. W., A. Bergman, D. D. Pollock and D. B. Goldstein. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* **145:** 207-216.

Felsenstein, J. 1993. PHYLIP: phylogeny inference package. Version 3.5c. University of Washington, Seattle.

Ferris, S. D., W. M. Brown, W. S. Davidson and A. C. Wilson. 1981. Extensive polymorphism in the mitochondrial DNA of apes. *Proc. Natl. Acad. Sci. USA* **78:** 6319-6323.

Foran, D. R., J. E. Hixson and W. M. Brown. 1988. Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nuc. Acids Res.* **16:** 5841-5861.

Fougerousse, F., R. Meloni, C. Roudaut and J. S. Beckmann. 1992. Tetranucleotide repeat polymorphism at the human N-MYC gene (MYCN). *Nuc. Acids Res.* **20**: 1165.

Frayer, D. W., M. H. Wolpoff, A. G. Thorne, F. H. Smith and G. G. Pope. 1993. Theories of modern human origins: the paleontological test. *Am. Anthropol.* **95:** 14-50.

Fu, Y.-X. and W.-H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133:** 693-709.

Fu, Y.-X. and W.-H. Li. 1996. Estimating the age of the common ancestor of men from the *ZFY* intron. *Science* **272:** 1356-1357.

Gabunia, L. and A. Vekua. 1995. A Plio-Pleistocene hominid from Dmanisi, East Georgia, Caucasus. *Nature* **373:** 509-512.

Gamble, C. S. 1994. *Timewalkers. The prehistory of global colonization.* Harvard University Press, Cambridge.

Garner, K. J. and O. A. Ryder. 1996. Mitochondrial DNA diversity in gorillas. *Mol. Phylogenet. Evol.* **6:** 39-48.

Garza, J. C., M. Slatkin and N. B. Freimer. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **12:** 594-603.

Gibbs, R. A., P. N. Nguyen, N. J. McBride, S. M. Koepf and C. T. Caskey. 1989. Identification of mutations leading to the Lesch-Nyhan syndrome by automated direct DNA sequencing of in vitro amplified DNA. *Proc. Natl. Acad. Sci. USA* **78:** 1919-1923.

Giles, R. E., H. Blanc, H. M. Cann and D. C. Wallace. 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **77:** 6715-6719.

Goldberg, T. L. and M. Ruvolo. 1997. The geographic apportionment of mitochondrial genetic diversity in east African chimpanzees, *Pan troglodytes schweinfurthii. Mol. Biol. Evol.* **14:** 976-984.

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza and M. W. Feldman. 1995*a*. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA* **92:** 6723-6727.

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza and M. W. Feldman. 1995*b*. An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139:** 463-471.

Gonder, M. K., J. F. Oates, T. R. Disotell, M. R. J. Forstner, J. C. Morales and D. J. Melnick. 1997. A new west African chimpanzee subspecies? *Nature* **388:** 337.

Graven, L., G. Passarino, O. Semino, P. Boursot, S. Santachiara-Benerecetti, A. Langaney and L. Excoffier. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol. Biol. Evol.* **12:** 334-345.

Greenberg, B. D., J. E. Newbold and A. Sugino. 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* **21:** 33-49.

Grunebaum, L., J. Cazenave, G. Camerino, C. Kloepfer, J. Mandel, P. Tolstoshev, M. Jaye, H. De la Salle and J. Lecocq. 1984. Carrier detection of hemophilia B by using a restriction site polymorphism associated with the coagulation Factor IX gene. *J. Clin. Invest.* **73:** 1491-1495.

Gyllensten, U. B. 1989. PCR and DNA sequencing. *Biotechniques* **7:** 700-708.

Gyllensten, U. B. and H. A. Erlich. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-*DQA* locus. *Proc. Natl. Acad. Sci. USA* **85:** 7652-7656.

Gyllensten, U. B., D. Wharton and A. C. Wilson. 1985. Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. *J. Hered.* **76:** 321-324.

Gyllensten, U. B., M. Sundvall and H. A. Erlich. 1991. Allelic diversity is generated by intraexon sequence exchange at the *DRB1* locus of primates. *Proc. Natl. Acad. Sci. USA* **88:** 3686-3690.

Gyllensten, U. B., T. Bergstrom, A. Josefsson, M. Sundvall and H. A. Erlich. 1996. Rapid allelic diversification and intensified selection at antigen recognition sites of the Mhc class II *DPB1* locus during hominoid evolution. *Tissue Ant.* **47:** 212-221.

Hale, L. R. and R. S. Singh. 1991. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history *vs.* selection in the genetic structure of geographic populations. *Genetics* **129:** 103-117.

Hammer, M. F. 1994. A recent insertion of an *Alu* element on the Y chromosome is a useful marker for human population studies. *Mol. Biol. Evol.* **11:** 749-761.

Hammer, M. F. 1995. A recent common ancestry for human Y chromosomes. *Nature* **378:** 376-378.

Hammer, M. F. and S. L. Zegura. 1996. The role of the Y chromosome in human evolutionary studies. *Evol. Anthropol.* **5:** 116-134.

Hammer, M. F., A. B. Spurdle, T. Karafet *et al.* 1997. The geographic distribution of human Y chromosome variation. *Genetics* **145:** 787-805.

Harding, R. M., S. M. Fullerton, R. C. Griffiths, J. Bond, M. J. Cox, J. A. Schneider, D. S. Moulin and J. B. Clegg. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* **60:** 772-789.

Harihara, S., N. Saitou, M. Hirai, N. Aoto, K. Terao, F. Cho, S. Honjo and K. Omoto. 1988. Differentiation of mitochondrial DNA types in *Macaca fascicularis. Primates* **29:** 117-127.

Harpending, H. C., S. T. Sherry, A. R. Rogers and M. Stoneking. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.* **34:** 483-496.

Harpending, H. C., J. H. Relethford and S. T. Sherry. 1996. Methods and models for understanding human diversity. Pp. 283-299 in A. J. Boyce and C. G. N. Mascie-Taylor (eds). *Molecular biology and human diversity*. Cambridge University Press, Cambridge.

Harris, H. and D. A. Hopkinson. 1972. Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphisms. *Ann. Hum. Genet.* **36:** 9-20.

Hasegawa, M., A. Di Rienzo, T. D. Kocher and A. C. Wilson. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. *J. Mol. Evol.* **37:** 347-354.

Hasegawa, M. and S. Horai. 1991. Time of the deepest root for polymorphism in human mitochondrial DNA. *J. Mol. Evol.* **32:** 37-42.

Hayasaka, K., S. Horai, T. Shotake, K. Nozawa and E. Matsunaga. 1986. Mitochondrial DNA polymorphism in Japanese monkeys, *Macaca fuscata. Jpn. J. Genet.* **61:** 345-359.

Hedges, S. B., S. Kumar, K. Tamura and M. Stoneking. 1992. Human origins and analysis of mitochondrial DNA sequences. *Science* **255:** 737-739.

Hey, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. *Mol. Biol. Evol.* **14:** 166-172.

Holland, M. M., D. L. Fisher, L. G. Mitchell, W. C. Rodriquez, J. J. Canik, C. R. Merril and V. W. Weedn. 1993. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the vietnam war. *J. Foren. Sci.* **38:** 542-553.

Horai, S. and K. Hayasaka. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* **46:** 828-842.

Horai, S., Y. Satta, K. Hayasaka, R. Kondo, T. Inoue, T. Ishida, S. Hayashi and N. Takahata. 1992. Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* **35:** 32-43.

Horai, S., R. Kondo, Y. Nakagawa-Hattori, S. Hayashi, S. Sonoda and K. Tajima. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol. Biol. Evol.* **10:** 23-47.

Horai, S., K. Hayasaka, R. Kondo, K. Tsugane and N. Takahata. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* **92:** 532-536.

Howell, N., I. Kubacka and D. A. Mackey. 1996. How rapidly does the human mitochondrial genome evolve? *Am. J. Hum. Genet.* **59:** 501-509.

Howells, W. W. 1942. Fossil man and the origin of the races. *Am. Anthropol.* **44:** 182-193.

Howells, W. W. 1959. *Mankind in the making.* Doubleday, Garden City.

Howells, W. W. 1976. Explaining modern man: evolutionists versus migrationists. *J. Hum. Evol.* **5:** 477-495.

Howells, W. W. 1993. *Getting here. The story of human evolution.* Compass Press, Washington.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Evol. Biol.* **7:** 1-44.

Hudson, R. R. 1994. How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? *Proc. Natl. Acad. Sci. USA* **91:** 6815-6818.

Hudson, R. R. and N. L. Kaplan. 1994. Gene trees with background selection. Pp. 140-153 in B. Golding (ed). *Non-neutral evolution: theories and molecular data.* Chapman and Hall, New York.

Hudson, R. R., M. Kreitman and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116:** 153-159.

Hudson, T. J., M. Engelstein, M. K. Lee, E. C. Ho, M. J. Rubenfield, C. P. Adams, D. E. Housman and N. C. Dracopoli. 1992. Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* **13:** 622-629.

Hutchison, C. A., J. E. Newbold, S. S. Potter and M. H. Edgell. 1974. Maternal inheritance of mitochondrial DNA. *Nature* **251:** 536-538.

Jakubiczka, S., J. Arnemann, H. J. Cooke, M. Krawczak and J. Schmidtke. 1989. A search for restriction fragment length polymorphism on the human Y chromosome. *Hum. Genet.* **84:** 86-88.

Jeffreys, A. J., V. Wilson and S. L. Thein. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314:** 67-73.

Jobling, M. A. and C. Tyler-Smith. 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* **11:** 449-456.

Johnson, M. J., D. C. Wallace, S. D. Ferris, M. C. Rattazzi and L. L. Cavalli-Sforza. 1983. Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.* **19:** 255-271.

Jorde, L. B., M. J. Bamshad, W. S. Watkins, R. Zenger, A. E. Fraley, P. A. Krakowiak, K. D. Carpenter, H. Soodyall, T. Jenkins and A. R. Rogers. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57:** 523-538.

Jorde, L. B., A. R. Rogers, M. Bamshad, W. S. Watkins, P. Krakowiak, S. Sung, J. Kere and H. C. Harpending. 1997. Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci. USA* **94:** 3100-3103.

Kaplan, N. L., R. R. Hudson and C. H. Langley. 1989. The "hitch-hiking effect" revisited. *Genetics* **123:** 887-899.

Kenter, M., N. Otting, J. Anholts, M. Jonker, R. Schipper and R. E. Bontrop. 1992. Mhc-*DRB* diversity of the chimpanzee (*Pan troglodytes*). *Immunogenetics* **37:** 1-11.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111-120.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

King, M. and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188:** 107-116.

Klein, J., N. Takahata and F. J. Ayala. 1993. MHC polymorphism and human origins. *Sci. Am.* **269:** 46-51.

Kocher, T. D. and A. C. Wilson. 1991. Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. Pp. 391-413 in S. Osawa and T. Honjo (eds). *Evolution of life: fossils, molecules and culture.* Springer-Verlag, Tokyo.

Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking and S. Pääbo. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* **90:** 19-30.

Krüger, J. and F. Vogel. 1989. The problem of our common mitochondrial mother. *Hum. Genet.* **82:** 308-312.

Kumar, S., K. Tamura and M. Nei. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.0. The Pennsylvania State University, University Park.

Kunkel, T. A. and L. A. Loeb. 1981. Fidelity of mammalian DNA polymerases. *Science* **213:** 765-767.

Lahr, M. M. and R. Foley. 1995. Multiple dispersals and modern human origins. *Evol. Anthropol.* **3:** 48-60.

Leakey, M. G., C. S. Feibel, I. McDougall and A. Walker. 1995. New four-million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature* **376:** 565-571.

Lewin, R. 1993. *The origin of modern humans*. Scientific American Library, New York.

Li, W.-H. 1978. Maintenance of genetic variability under the joint effect of mutation, selection and random drift. *Genetics* **90:** 349-382.

Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36:** 96-99.

Linares, A. R., K. Nayar, D. B. Goldstein, J. M. Hebert, M. T. Seielstad, P. A. Underhill, A. A. Lin, M. W. Feldman and L. L. Cavalli-Sforza. 1996. Geographic clustering of human Y-chromosome haplotypes. *Ann. Hum. Genet.* **60:** 401-408.

Litt, M. and J. A. Luty. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44:** 397-401.

Loewe, L. and S. Scherer. 1997. Mitochondrial Eve: the plot thickens. *Trends Ecol. Evol.* **12:** 422-423.

Long, J. C., A. Chakravarti, C. D. Boehm, S. Antonarakis and H. H. Kazazian. 1990. Phylogeny of human β-globin haplotypes and its implications for recent human evolution. *Am. J. Phys. Anthropol.* **81:** 113-130.

Lu, J., R. Riley, M. Robertson, L. Nelson and K. Ward. 1993. Tetranucleotide repeat polymorphisms at the D8S342, D8S323, D8S345, D8S315 and D8S347 loci on 8q. *Hum. Mol. Genet.* **2:** 1743.

Lucotte, G., P. Guérin, L. Hallé, F. Loirat and S. Hazout. 1989. Y chromosome DNA polymorphisms in two African populations. *Am. J. Hum. Genet.* **45:** 16-20.

Maddison, D. R. 1991. African origin of human mitochondrial DNA reexamined. *Syst. Zool.* **40:** 355-363.

Maddison, D. R., M. Ruvolo and D. L. Swofford. 1992. Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. *Syst. Biol.* **41:** 111-124.

Maeda, N., J. B. Bliska and O. Smithies. 1983. Recombination and balanced chromosome polymorphism suggested by DNA sequences 5' to the human δ-globin gene. *Proc. Natl. Acad. Sci. USA* **80:** 5012-5016.

Malaspina, P., F. Persichetti, A. Novelletto, C. Iodice, L. Terrenato, J. Wolfe, M. Ferraro and G. Prantera. 1990. The human Y chromosome shows a low level of DNA polymorphism. *Ann. Hum. Genet.* **54:** 297-305.

Maniatis, T., E. F. Fritsch and J. Sambrook. 1982. *Molecular Cloning (A Laboratory Manual).* Cold Spring Harbor Laboratory, New York.

Mathias, N., M. Bayés and C. Tyler-Smith. 1994. Highly informative compound haplotypes for the human Y chromosome. *Hum. Mol. Genet.* **3:** 115-123.

Mayer, W. E., C. O'hUigin, Z. Zaleska-Rutczynska and J. Klein. 1992. Trans-species origin of Mhc-*DRB* polymorphism in the chimpanzee. *Immunogenetics* **37:** 12-23.

Maynard-Smith, J. 1994. Estimating selection by comparing synonymous and substitutional changes. *J. Mol. Evol.* **39:** 123-128.

Maynard-Smith, J. and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23:** 23-35.

McDonald, J. H. and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila. Nature* **351:** 652-654.

Merriwether, D. A., A. G. Clark, S. W. Ballinger, T. G. Schurr, H. Soodyall, T. Jenkins, S. T. Sherry and D. C. Wallace. 1991. The structure of human mitochondrial DNA variation. *J. Mol. Evol.* **33:** 543-555.

Michalakis, Y. and L. A. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142:** 1061-1064.

Miller, S. A., D. D. Dykes and H. F. Polesky. 1988. A simple salting-out procedure for extracting DNA from human nucleated cells. *Nuc. Acids Res.* **16:** 1215.

Miquel, J. 1992. An update on the mitochondrial DNA mutation hypothesis of cell aging. *Mutat. Res.* **275:** 209-216.

Mitchell, R. J. and M. F. Hammer. 1996. Human evolution and the Y chromosome. *Curr. Opin. Genet. Dev.* **6:** 737-742.

Miyamoto, M. M., J. L. Slightom and M. Goodman. 1987. Phylogenetic relations of humans and African apes from DNA sequences in the ψη-globin region. *Science* **238:** 369-373.

Morin, P. A., J. J. Moore, R. Chakraborty, L. Jin, J. Goodall and D. S. Woodruff. 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265:** 1193-1201.

Mountain, J. L. and L. L. Cavalli-Sforza. 1994. Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. *Proc. Natl. Acad. Sci. USA* **91:** 6515-6519.

Mountain, J. L., J. M. Hebert, S. Bhattacharyya, P. A. Underhill, C. Ottolenghi, M. Gadgil and L. L. Cavalli-Sforza. 1995. Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* **56:** 979-992.

Nachman, M. W., S. N. Boyer and C. F. Aquadro. 1994. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. *Proc. Natl. Acad. Sci. USA* **91:** 6364-6368.

Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* **142:** 953-963.

Nakamura, Y., M. Leppert, P. O'Connell *et al.* 1987. Variable number tandem repeat (VNTR) markers for human gene mapping. *Science* **235:** 1616-1622.

Nauta, M. J. and F. J. Weissing. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143:** 1021-1032.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583-590.

Nei, M. 1985. Human evolution at the molecular level. Pp. 41-64 in T. Ohta and K. Aoki (eds). *Population genetics and molecular evolution.* Springer-Verlag, Berlin.

Nei, M. 1987. *Molecular evolutionary genetics.* Columbia University Press, New York.

Nei, M. 1992. Age of the common ancestor of human mitochondrial DNA. *Mol. Biol. Evol.* **9:** 1176-1178.

Nei, M. and D. Graur. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* **17:** 73-118.

Nei, M. and G. Livshits. 1989. Genetic relationships of Europeans, Asians and Africans and the origin of modern *Homo sapiens. Hum. Hered.* **39:** 276-281.

Nei, M. and T. Ota. 1991. Evolutionary relationships of human populations at the molecular level. Pp. 415-428 in S. Osawa and T. Homjo (eds). *Evolution of life.* Springer-Verlag, Tokyo.

Nei, M. and A. K. Roychoudhury. 1974*a*. Genic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids. *Am. J. Hum. Genet.* **26:** 421-443.

Nei, M. and A. K. Roychoudhury. 1974*b*. Sampling variances of heterozygosity and genetic distance. *Genetics* **76:** 379-390.

Nei, M. and A. K. Roychoudhury. 1982. Genetic relationship and evolution of human races. *Evol. Biol.* **14:** 1-59.

Nei, M. and A. K. Roychoudhury. 1993. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* **10:** 927-943.

Nei, M. and N. Takezaki. 1996. The root of the phylogenetic tree of human populations. *Mol. Biol. Evol.* **13:** 170-177.

Noble, W. and I. Davidson. 1991. The evolutionary emergence of modern human behavior: language and its archaeology. *Man* **26:** 223-254.

Nozawa, K., T. Shotake, Y. Kawamoto and Y. Tanabe. 1982. Population genetics of Japanese monkeys: II. Blood protein polymorphisms and population structure. *Primates* **23:** 252-271.

Oakey, R. and C. Tyler-Smith. 1990. Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. *Genomics* **7:** 325-330.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23:** 263-286.

Ohta, T. and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* **22:** 201-204.

Parsons, T. J. and M. M. Holland. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nat. Genet.* **18:** 110.

Parsons, T. J., D. S. Muniec, K. Sullivan *et al.* 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nat. Genet.* **15:** 363-367.

Pascall, V. L., G. Destro Bisol, M. Dobosz, E. d'Aloja, G. Paonessa and U. Mereu. 1994. Chimpanzee DNA profiles on trial. *Nature* **367:** 692-693.

Pääbo, S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. USA* **86**: 1939-1943.

Penny, D., M. Steel, P. J. Waddell and M. D. Hendy. 1995. Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Mol. Biol. Evol.* **12:** 863-882.

Perna, N. T., M. A. Batzer, P. L. Deininger and M. Stoneking. 1992. *Alu* insertion polymorphism: a new type of marker for human population studies. *Hum. Biol.* **64:** 641-648.

Persichetti, F., P. Blasi, M. Hammer, P. Malasapina, C. Jodice, L. Terrenato and A. Novelletto. 1992. Disequilibrium of multiple DNA markers on the human Y chromosome. *Ann. Hum. Genet.* **56:** 303-310.

Persson, B. and P. Argos. 1994. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.* **237:** 182-192.

Pesole, G., E. Sbisa, G. Preparata and C. Saccone. 1992. The evolution of the mitochondrial D-loop region and the origin of modern man. *Mol. Biol. Evol.* **9:** 587-598.

Polymeropoulos, M. H., H. Xiao, D. S. Rath and C. R. Merril. 1991. Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nuc. Acids Res.* **19:** 3753.

Pritchard, J. K. and M. W. Feldman. 1996. Genetic data and the African origin of humans. *Science* **274:** 1548-1549.

Protsch, R. 1975. The absolute dating of Upper Pleistocene sub-Saharan fossil hominids and their place in human evolution. *J. Hum. Evol.* **4:** 297-322.

Rand, D. M., M. Dorfsman and L. M. Kann. 1994. Neutral and non-neutral evolution of Drosophila mitochondrial DNA. *Genetics* **138:** 741-756.

Rapacz, J., L. Chen, E. Butler-Brunner, M.-J. Wu, J. O. Hasler-Rapacz, R. Butler and V. N. Schumaker. 1991. Identification of the ancestral haplotype for apolipoprotein B suggests an African origin of *Homo sapiens sapiens* and traces their subsequent migration to Europe and the Pacific. *Proc. Natl. Acad. Sci. USA* **88:** 1403-1406.

Relethford, J. H. 1995. Genetics and modern human origins. *Evol. Anthropol.* **4:** 53-63.

Relethford, J. H. 1997. Mutation rate and excess African heterozygosity. *Hum. Biol.* **69:** 785-792.

Relethford, J. H. and H. C. Harpending. 1994. Craniometric variation, genetic theory, and modern human origins. *Am. J. Phys. Anthropol.* **95:** 249-270.

Relethford, J. H. and H. C. Harpending. 1995. Ancient differences in population size can mimic a recent African origin of modern humans. *Curr. Anthropol.* **36:** 667-674.

Richards, R. I., J. Crawford, K. Narahara *et al.* 1996. Dynamic mutation loci: allele distributions in different populations. *Ann. Hum. Genet.* **60:** 391-400.

Rogers, A. R. 1995. Genetic evidence for a pleistocene population explosion. *Evolution* **49:** 608-615.

Rogers, A. R. and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9:** 552-569.

Rogers, A. R. and L. B. Jorde. 1995. Genetic evidence on modern human origins. *Hum. Biol.* **67:** 1-36.

Rogers, A. R. and L. B. Jorde. 1996. Ascertainment bias in estimates of average heterozygosity. *Am. J. Hum. Genet.* **58:** 1033-1041.

Rogers, J., P. B. Samollow and A. G. Comuzzie. 1996. Estimating the age of the common ancestor of men from the *ZFY* intron. *Science* **272:** 1360-1361.

Rubinsztein, D. C., W. Amos, J. Leggo, S. Goodburn, S. Jain, S. Li, R. L. Margolis, C. A. Ross and M. A. Ferguson-Smith. 1995*a*. Microsatellite evolution - evidence for directionality and variation in rate between species. *Nat. Genet.* **10:** 337-343.

Rubinsztein, D. C., J. Leggo and W. Amos. 1995*b*. Microsatellites evolve more rapidly in humans than in chimpanzees. *Genomics* **30:** 610-612.

Ruvolo, M. 1996. A new approach to studying modern human origins: hypothesis testing with coalescence time distributions. *Mol. Phylogenet. Evol.* **5:** 202-219.

Ruvolo, M., S. Zehr, M. von Dornum, D. Pan, B. Chang and J. Lin. 1993. Mitochondrial COII sequences and modern human origins. *Mol. Biol. Evol.* **10:** 1115-1135.

Ruvolo, M., D. Pan, S. Zehr, T. Goldberg, T. R. Disotell and M. von Dornum. 1994. Gene trees and hominoid phylogeny. *Proc. Natl. Acad. Sci. USA* **91:** 8900-8904.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487-491.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406-425.

Saitou, N. and K. Omoto. 1987. Time and place of human origins from mtDNA data. *Nature* **327:** 288.

Savatier, P., G. Trabuchet, C. Faure, Y. Chebloune, M. Gouy, G. Verdier and V. M. Nigon. 1985. Evolution of the primate β-globin gene region. High rate of variation in CpG dinucleotides and in short repeated sequences between man and chimpanzee. *J. Mol. Biol.* **182:** 21-29.

Seielstad, M. T., J. M. Hebert, A. A. Lin, P. A. Underhill, M. Ibrahim, D. Vollrath and L. L. Cavalli-Sforza. 1994. Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.* **3:** 2159-2161.

Seigel, S. 1956. *Nonparametric statistics for the behavioural sciences.* McGraw Hill, Tokyo.

Sherry, S. T., A. R. Rogers, H. Harpending, H. Soodyall, T. Jenkins and M. Stoneking. 1994. Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* **66:** 761-775.

Shoffner, J. M. and D. C. Wallace. 1992. Mitochondrial genetics: principles and practice. *Am. J. Hum. Genet.* **51:** 1179-1186.

Shriver, M. D., L. Jin, R. Chakraborty and E. Boerwinkle. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* **134:** 983-993.

Shriver, M. D., L. Jin, E. Boerwinkle, R. Deka, R. E. Ferrell and R. Chakraborty. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* **12:** 914-920.

Shriver, M. D., L. Jin, R. E. Ferrell and R. Deka. 1997. Microsatellite data support an early population expansion in Africa. *Genome Res.* **7:** 586-591.

Sibley, C. G. and J. E. Ahlquist. 1987. DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. *J. Mol. Evol.* **26:** 99-121.

Simonsen, K. L., G. A. Churchill and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141:** 413-429.

Slagel, V., E. Flemington, V. Traina-Dorge, H. Bradshaw and P. Deininger. 1987. Clustering and subfamily relationships of the Alu family in the human genome. *Mol. Biol. Evol.* **4**: 19-29.

Slatkin, M. 1985. Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* **16:** 393-430.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139:** 457-462.

Slatkin, M. and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129:** 555-562.

Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent and L. E. Hood. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321:** 674-679.

Smith, S. W., R. Overbeek, C. R. Woese, W. Gilbert and P. M. Gillevet. 1994. The genetic data environment and expandable GUI for multiple sequence analysis. *Comp. Appl. Biosci.* **10:** 671-675.

Sokal, R. R. and F. J. Rohlf. 1995. *Biometry: the principles and practice of statistics in biological research.* 3rd edition. W. H. Freeman and Company, New York.

Spurdle, A. B. and T. Jenkins. 1991. Y chromosome probe p49a detects complex *Pvu*II haplotypes and many new *Taq*I haplotypes in Southern African populations. *Am. J. Hum. Genet.* **50:** 107-125.

Spurdle, A. B. and T. Jenkins. 1992. The search for Y chromosome polymorphism is extended to negroids. *Hum. Mol. Genet.* **1:** 169-170.

Spurdle, A. B., M. F. Hammer and T. Jenkins. 1994. The Y *Alu* polymorphism in Southern African populations and its relationship to other Y-specific polymorphisms. *Am. J. Hum. Genet.* **54:** 319-330.

Stephan, W., T. H. E. Wiehe and M. W. Lenz. 1992. The effect of strongly selected substitutions on neutral polymorphism–analytical results based on diffusion theory. *Theor. Pop. Biol.* **41:** 237-254.

Stoneking, M. 1993. DNA and recent human evolution. *Evol. Anthropol.* **2:** 60-73.

Stoneking, M., K. Bhatia and A. C. Wilson. 1986. Rate of sequence divergence estimated from restriction maps of mitochondrial DNAs from Papua New Guinea. *Cold Spring Harb. Symp. Quant. Biol.* **LI:** 433-438.

Stoneking, M., D. Hedgecock, R. G. Higuchi, L. Vigilant and H. A. Erlich. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.* **48:** 370-382.

Stoneking, M., S. T. Sherry, A. J. Redd and L. Vigilant. 1992. New approaches to dating suggest a recent age for the human mtDNA ancestor. *Phil. Trans. R. Soc. Lond. B* **337:** 167-175.

Stoneking, M., J. J. Fontius, S. L. Clifford, H. Soodyall, S. S. Arcot, N. Saha, T. Jenkins, M. A. Tahir, P. L. Deininger and M. A. Batzer. 1997. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7:** 1061-1071.

Strimmer, K. and A. von Haeseler. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA* **94:** 6815-6818.

Stringer, C. B. 1990. The emergence of modern humans. *Sci. Am.* **263:** 68-74.

Stringer, C. B. and P. Andrews. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* **239:** 1263-1268.

Suggs, S. V., T. Hirose, T. Miyake, E. H. Kawashima, M. J. Johnson, K. Itakura and R. B. Wallace. 1981. *Developmental biology using purified genes.* Academic Press, New York.

Sullivan, K. M., R. Hopgood and P. Gill. 1992. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int. J. Leg. Med.* **105:** 83-86.

Swisher, C. C., G. H. Curtis, T. Jacob, A. G. Getty, A. Suprijo and Widiasmoro. 1994. Age of the earliest known hominids in Java, Indonesia. *Science* **263:** 1118-1121.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105:** 437-460.

Tajima, F. 1989*a*. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585-595.

Tajima, F. 1989*b*. The effect of change in population size on DNA polymorphism. *Genetics* **123:** 597-601.

Tajima, F. 1993. Measurement of DNA polymorphism. Pp. 37-59 in N. Takahata and A. G. Clark (eds). *Mechanisms of molecular evolution.* Japan Scientific Societies Press, Tokyo.

Takahata, N. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. Natl. Acad. Sci. USA* **87:** 2419-2423.

Takahata, N. 1991. Trans-species polymorphism of HLA molecules, founder principle, and human evolution. Pp. 29-49 in J. Klein and D. Klein (eds). *Molecular evolution of the major histocompatibility complex.* Springer-Verlag, Berlin.

Takahata, N. 1993*a*. Allelic genealogy and human evolution. *Mol. Biol. Evol.* **10:** 2-22.

Takahata, N. 1993*b*. Relaxed natural selection in human populations during the Pleistocene. *Jpn. J. Genet.* **68:** 539-547.

Tamura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10:** 512-526.

Tavaré, S., D. J. Balding, R. C. Griffiths and P. Donnelly. 1997. Inferring coalescence times from DNA sequence data. *Genetics* **145:** 505-518.

Templeton, A. R. 1987. Genetic systems and evolutionary rates. Pp. 218-234 in K. S. W. Campbell and M. F. Day (eds). *Rates of Evolution.* Allen and Unwin, London.

Templeton, A. R. 1992. Human origins and analysis of mitochondrial DNA sequences. *Science* **255:** 737.

Templeton, A. R. 1993. The "Eve" hypotheses: a genetic critique and reanalysis. *Am. Anthropol.* **95:** 51-72.

Templeton, A. R. 1994. "Eve": hypothesis compatibility vs. hypothesis testing. *Am. Anthropol.* **96:** 141-147.

Templeton, A. R. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* **144:** 1263-1270.

Thyagarajan, B., R. A. Padua and C. Campbell. 1996. Mammalian mitochondria possess homologous DNA recombination activity. *J. Biol. Chem.* **271:** 27536-27543.

Tishkoff, S. A., E. Dietzsch, W. Speed *et al.* 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271:** 1380-1387.

Titus-Trachtenberg, E. A., O. Rickards, G. F. De Stefano and H. A. Erlich. 1994. Analysis of HLA class II haplotypes in the Cayapa Indians of Ecuador: a novel *DRB1* allele reveals evidence for convergent evolution and balancing selection at position 86. *Am. J. Hum. Genet.* **55:** 160-167.

Torroni, A., O. Semino, R. Scozzari, G. Sirugo, G. Spedini, N. Abbas, M. Fellous and A. S. S. Benerecetti. 1990. Y chromosome DNA polymorphisms in human populations: differences between Caucasoids and Africans detected by 49a and 49f probes. *Ann. Hum. Genet.* **54:** 287-296.

Underhill, P. A., L. Jin, R. Zemans, P. J. Oefner and L. L. Cavalli-Sforza. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. USA* **93:** 196-200.

Underhill, P. A., L. Jin, A. A. Lin, S. Q. Mehdi, T. Jenkins, D. Vollrath, R. W. Davis, L. L. Cavalli-Sforza and P. J. Oefner. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7:** 996-1005.

Valdes, A. M., M. Slatkin and N. B. Freimer. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133:** 737-749.

Vigilant, L., R. Pennington, H. Harpending, T. D. Kocher and A. C. Wilson. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci. USA* **86:** 9350-9354.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A. C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253:** 1503-1507.

von Haeseler, A., A. Sajantila and S. Pääbo. 1995. The genetical archaeology of the human genome. *Nat. Genet.* **14:** 135-140.

Wainscoat, J. S., A. V. S. Hill, A. L. Boyce *et al.* 1986. Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* **319**: 491-493.

Wakeley, J. 1993. Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. *J. Mol. Evol.* **37:** 613-623.

Wallace, D. C. 1992. Diseases of the mitochondrial DNA. *Annu. Rev. Biochem.* **61:** 1175-1212.

Wallace, D. C. 1995. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am. J. Hum. Genet.* **57:** 201-223.

Ward, R. H., B. L. Frazier, K. Dew-Jager and S. Pääbo. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* **88:** 8720-8724.

Watanabe, T., M. Mizutani, S. Wakana and T. Tomita. 1985. Demonstration of the maternal inheritance of avian mitochondrial DNA in chicken-quail hybrids. *J. Exp. Zool.* **236:** 245-247.

Watson, E. 1996. Threads from the past: a genetic study of African ethnic groups and human origins. Thesis, Massey University.

Watson, E., K. Bauer, R. Aman, G. Weiss, A. von Haeseler and S. Pääbo. 1996. mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.* **59:** 437-444.

Watterson, G. A. 1975. On the number of segregating sites in genetic models without recombination. *Theor. Pop. Biol.* **7:** 256-276.

Watterson, G. A. 1978. The homozygosity test of neutrality. *Genetics* **88:** 405-417.

Weber, J. L. and C. Wong. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2:** 1123-1128.

Weidenreich, F. 1940. Some problems dealing with ancient man. *Am. Anthropol.* **42:** 375-383.

Weidenreich, F. 1946. *Apes, Giants, and Man.* University of Chicago Press, Chicago.

Weiss, K. M. 1988. In search of times past: Gene flow and invasion in the generation of human diversity. Pp. 130-166 in C. G. Mascie-Taylor and G. W. Lasker (eds). *Biological aspects of human migration.* Cambridge University Press, Cambridge.

Weiss, K. M. and T. Maruyama. 1976. Archeology, population genetics and studies of human racial ancestry. *Am. J. Phys. Anthropol.* **44:** 31-49.

Weiss, G. and A. von Haeseler. 1996. Estimating the age of the common ancestor of men from the *ZFY* intron. *Science* **272:** 1359-1360.

White, T. D., G. Suwa and B. Asfaw. 1994. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **371:** 306-312.

White, T. D., G. Suwa and B. Asfaw. 1995. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **375:** 88.

Whitfield, L. S., J. E. Sulston and P. N. Goodfellow. 1995. Sequence variation of the human Y chromosome. *Nature* **378:** 379-380.

Whittam, T. S., A. G. Clark, M. Stoneking, R. L. Cann and A. C. Wilson. 1986. Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. *Proc. Natl. Acad. Sci. USA* **83:** 9611-9615.

Willard, C., H. T. Nguyen and C. W. Schmid. 1987. Existence of at least three distinct Alu subfamilies. *J. Mol. Evol.* **26:** 180-186.

Wills, C. 1995. When did eve live? An evolutionary detective story. *Evolution* **49:** 593-607.

Wilson, A. C., R. L. Cann, S. M. Carr *et al.* 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* **26:** 375-400.

Wise, C. A., M. Sraml, D. C. Rubinsztein and S. Easteal. 1997. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. *Mol. Biol. Evol.* **14:** 707-716.

Wise, C. A., M. Sraml and S. Easteal. 1998. Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. *Genetics* **148:** 409-421.

WoldeGabriel, G., T. D. White, G. Suwa, P. Renne, J. de Heinzelin, W. K. Hart and G. Heiken. 1994. Ecological and temporal placement of early Pliocene hominids at Aramis, Ethiopia. *Nature* **371:** 330-333.

Wolff, R., Y. Nakamura, S. Odelberg, R. Shiang and R. White. 1991. Generation of variability at VNTR loci in human DNA. Pp. 20-38 in T. Burke, G. Dolf, A. J. Jeffreys and R. Wolff (eds). *DNA fingerprinting: approaches and applications.* Birkhauser Verlag, Basel.

Wolpoff, M. H. 1989. Multiregional evolution: the fossil alternative to Eden. Pp. 62-108 in P. Mellars and C. Stringer (eds). *The human revolution: behavioural and biological perspectives on the origins of modern humans.* Princeton University Press, New Jersey.

Wolpoff, M. H. and R. Caspari. 1997. *Race and human evolution.* Simon and Schuster, New York.

Wolpoff, M. H., W. Xinzhi and A. G. Thorne. 1984. Modern Homo sapiens origins: a general theory of hominid evolution involving the fossil evidence from east Asia. Pp. 411-483 in F. H. Smith and F. Spencer (eds). *The origins of modern humans: a world survey of the fossil evidence.* Alan R. Liss, New York.

Wright, S. 1969. *Evolution and the genetics of populations: the theory of gene frequencies*. University of Chicago Press, Chicago.

Wrischnik, L. A., R. G. Higuchi, M. Stoneking, H. A. Erlich, N. Arnheim and A. C. Wilson. 1987. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nuc. Acids. Res.* **15:** 529-542.

Wu, S., S. Seino and G. I. Bell. 1990. Human collagen, type II, alpha 1, (COL2A1) gene: VNTR polymorphism detected by gene amplification. *Nuc. Acids Res.* **18:** 3102.

Xiong, W., W.-H. Li, I. Posner, T. Yamamura, A. Yamamoto, A. M. Gotto and L. Chan. 1991. No severe bottleneck during human evolution: evidence from two apolipoprotein C-II deficiency alleles. *Am. J. Hum. Genet.* **48:** 383-389.

Xu, X. and U. Arnason. 1996. The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. *J. Mol. Evol.* **43:** 431-437.

Zaykin, D. V. and A. I. Pudovkin. 1993. Two programs to estimate significance of χ^2 values using pseudo-probability tests. *J. Hered.* **84:** 152.

Zerjal, T., B. Dashnyam, A. Pandya *et al.* 1997. Genetic relationships of Asians and northern Europeans, revealed by Y-chromosome DNA analysis. *Am. J. Hum. Genet.* **60:** 1174-1183.

Zhivotovsky, L. A. and M. W. Feldman. 1995. Microsatellite variability and genetic distances. *Proc. Natl. Acad. Sci. USA* **92:** 11549-11552.

Zhivotovsky, L. A., M. W. Feldman and S. A. Grishechkin. 1997. Biased mutations and microsatellite variation. *Mol. Biol. Evol.* **14**: 926-933.

Zischler, H., H. Geisert, A. von Haeseler and S. Pääbo. 1995. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **378:** 489-492.